



НАЦИОНАЛЕН ВОЕНЕН УНИВЕРСИТЕТ „ВАСИЛ ЛЕВСКИ”

ФАКУЛТЕТ „АРТИЛЕРИЯ, ПВО И КИС”

9713 гр. Шумен, ул. „Карел Шкорпил” №1

телефон: (054) 801 040; тел./факс: (054) 877 463;

e-mail: nvu-sh@aadcf.nvu.bg



Катедра „Компютърни системи и технологии“

асистент магистър физик Любомир Евгениев Манов

**ОТКРИВАНЕ И РАЗПОЗНАВАНЕ НА ЧАСТИЧНО ВИДИМИ
ОБЕКТИ С ВОЕННО ПРЕДНАЗНАЧЕНИЕ В ОПТИЧНИ
ИЗОБРАЖЕНИЯ С ИЗПОЛЗВАНЕ НА ИЗКУСТВЕН ИНТЕЛЕКТ**

АВТОРЕФЕРАТ

на дисертационен труд за присъждане на
образователна и научна степен „ДОКТОР“

Докторска програма по научна специалност:

„Автоматизирани системи за обработка на информация и управление“

НАУЧНА ОБЛАСТ: 5. ТЕХНИЧЕСКИ НАУКИ

ПРОФЕСИОНАЛНО НАПРАВЛЕНИЕ 5.3. КОМУНИКАЦИОННА И КОМПЮТЪРНА
ТЕХНИКА.

Научен ръководител:

Полк. проф. д-р инж. Чавдар Николаев Минчев

Шумен

2026 г.

Дисертационния труд се състои от 175 страници.

Брой на таблиците – 29.

Брой на фигурите – 48.

Брой на приложенията – 2.

Брой на литературните източници – 136.

Брой на публикациите по дисертацията – 5.

Защитата на дисертационния труд ще се състои на 11.05.2026 г. от ____.
ч. в зала _____ на _____.

Материалите по защитата са на разположение на интересуващите се в
_____ на _____, тел. _____.

Дисертационният труд е обсъден на катедрен съвет на катедра „Компютърни системи и технологии” при факултет „Артилерия, ПВО и КИС” на Национален военен университет „Васил Левски” – Велико Търново на 04.02.2026 г., и е насрочен за защита пред научно жури за придобиване на образователната и научна степен “Доктор”.

Докторантът е асистент в катедра „Радарни и комуникационни системи” при факултет „Артилерия, ПВО и КИС” на Национален военен университет „Васил Левски” – Велико Търново.

Основните изследвания по дисертационния труд са проведени в развойната програмна среда MATLAB.

Автор: асистент магистър-физик Любомир Евгениев Манов.

Тема: „Откриване и разпознаване на частично видими обекти с военно предназначение в оптични изображения с използване на изкуствен интелект”.

Тираж ____

Отпечатан на ____ . ____ . 20__ г.

Издателски комплекс на Национален военен университет „Васил Левски“

1. ОБЩА ХАРАКТЕРИСТИКА НА ДИСЕРТАЦИОННИЯ ТРУД.

Мотиви за разработка на дисертационния труд

Мотивите за разработка на дисертационния труд произтичат от нарастващите изисквания към съвременните военни системи за наблюдение и обективните ограничения на традиционните методи за обработка на оптични изображения. В реални условия обектите с военно и стратегическо значение често са частично закрити, маскирани или наблюдавани при неблагоприятни условия. Това води до ниско качество на изображенията и силно затруднява надеждното им откриване и разпознаване. Съществува осезаема потребност от методи и алгоритми, които да осигуряват по-висока точност и устойчивост при разпознаване на частично видими обекти в сложна обстановка.

Паралелно с това развитието на изкуствения интелект и дълбоко структурираното обучение на многослойни невронни мрежи през последното десетилетие разкрива нови възможности за автоматизирана обработка на големи обеми от визуална информация. Конволюционните невронни мрежи (CNN) и свързаните с тях архитектури показват значителен напредък при решаването на задачи по откриване и разпознаване на обекти в изображения, но приложението им в специфичния домейн на военните оптични изображения с частична видимост остава недостатъчно изследвано. Това е предпоставка за разработка и изследване на адаптирани към домейна, невронни архитектури, тренировъчни стратегии и набори от данни, съобразени с особеностите на задачата.

Допълнителен мотив за разработката на дисертационния труд е необходимостта от повишаване на степента на автоматизация и обективност при вземането на решения на тактическо и оперативно ниво. Чрез внедряване на методи на изкуствения интелект в системите за наблюдение и анализ се цели намаляване на натоварването върху операторите, ограничаване на субективния фактор и повишаване на съвременността и надеждността на предоставяната информация.

Актуалност на проблема

В условията на нарастваща заплаха от локални и асиметрични военни конфликти необходимостта от постоянна, надеждна и навременна информация за обекти с военно значение е от особена важност. Съвременните военни и разузнавателни системи генерират огромни обеми от данни от разнообразни източници – както военни, така и цивилни. Съществена част от тази информация постъпва под формата на оптични изображения, получени от разстояние от разнородни платформи за наблюдение – спътникови системи, пилотирувани и непилотирувани летателни апарати, наземни системи.

На практика, значителна част от тези изображения се характеризират с ниско качество и ограничена пространствена разделителна способност.

Получени са при неблагоприятни условия на наблюдение и частична видимост на обектите, поради закриване, маскировка или наличие на сложен фон. Това съществено затруднява традиционните методи за автоматизирана обработка и изисква разработването на по-устойчиви подходи за откриване и разпознаване.

За да се осигури ефективна обработка и интерпретация на този информационен поток, е необходимо внедряването на съвременни автоматизирани системи, базирани на методи на изкуствения интелект и в частност дълбоко структурираното обучение на многослойни невронни мрежи. Крайната цел е човешкият фактор, ангажиран с вземането на решения на локално, тактическо, оперативно и стратегическо ниво, да разполага с точна, надеждна и своєвременна информация, която да подпомага планирането и управлението на военните операции.

Цел и задачи на дисертационния труд

Основната цел на дисертационния труд е да се проектира и оптимизира многослойна невронна мрежа с дълбоко структурирано обучение за откриване и разпознаване на частично видими обекти с военно предназначение в оптични изображения.

Произтичащите от поставената цел **задачи** са:

1. Да се извърши анализ на съществуващите подходи за откриване и разпознаване на обекти в оптични изображения с използване на невронни мрежи с акцент върху приложенията в отбраната.

2. Да се структурира специализирана база от данни от оптични изображения на военна техника, включително сцени с реални бойни и тренировъчни условия, с частично закриване на обектите и наличие на множество обекти в кадъра.

3. Да се извърши систематична оценка на набор от съвременни предварително обучени невронни мрежи чрез създадената база от данни от оптични изображения на военна техника.

4. Да се осъществи избор на невронна архитектура на многослойна невронна мрежа за откриване и разпознаване на обекти в оптични изображения.

5. Да се проектира, реализира, оптимизира и оцени невронна мрежа за откриване, разпознаване и класификация на частично закрити военни обекти в оптични изображения.

Предмет на дисертационния труд

Предмет на дисертационния труд са модели и алгоритми, базирани на многослойни невронни мрежи с дълбоко структурирано обучение, за разпознаване на оптични изображения.

Средства за постигане на поставената цел

- методи за цифрова обработка на изображения;

- теория и принципи за изграждане и работа на системи с изкуствен интелект за разпознаване на изображения;
- теория на компютърното зрение;
- статистически методи за обработка на информация;
- методи за програмно и стимулационно моделиране в развойната софтуерна среда на Matlab/Simulink.

Научна новост

Научната новост в дисертационния труд се заключава в предложения подход за разпознаване на частично закрити военни обекти в оптични изображения от структурирана специализирана база от данни с използване на невронни архитектури с дълбоко структурирано обучение.

Практическа полезност

Практическата полезност се състои в разработване на математически модели и алгоритми за откриване и разпознаване на частично видими обекти с военно предназначение в оптични изображения с приложимост в съвременни системи за разузнаване, наблюдение и целеуказване, използвани от въоръжените сили и структури за сигурност, с цел повишаване на надеждността и бързината на оценка на обстановката на локално, тактическо и оперативно ниво.

Структура на дисертационния труд

В настоящия дисертационен труд се разработват и изследват подходи и архитектури за автоматично откриване и разпознаване на частично видими обекти с военно предназначение в оптични изображения, базирани на многослойни невронни мрежи с дълбоко структурирано обучение.

В **предговора** са формулирани мотивите и актуалността на изследването, предметът на научното изследване, целта и задачите на дисертационния труд, използваните методи на изследване, научните приноси и практическата полезност.

Дисертационният труд съдържа предговор, четири глави, заключение, списък на научните приноси, списък на публикациите, списък на цитираната литература и приложения. Четирите глави са логически обвързани и следват естествената последователност от анализ на съществуващото състояние на проблема, теоретично обосноваване на използваните методи и модели, провеждане на систематични компютърни експерименти и формулиране на конкретни научни приноси.

Първа глава представя систематичен литературен обзор на съвременното състояние на методите за автоматично откриване и разпознаване на обекти в оптични изображения. Анализирани са класическите подходи, базирани на ръчно дефинирани признаци, съвременните дълбоки невронни архитектури и специфичните предизвикателства, свързани с

частичната видимост, оклузиите и ограничените обучаващи данни в специализирани домейни. Обзорът обхваща 68 източника и служи като основа за формулиране на подход към решаване на поставените задачи на дисертацията.

Втора глава излага теоретичните основи и математическите модели на многослойните невронни мрежи с дълбоко структурирано обучение. Разгледани са архитектурата и принципите на работа на CNN, алгоритмите за обучение и регуларизация, функциите на загубата и основните свързани архитектури. В края на главата е формализиран математически модел за класификация на частично видими обекти чрез въвеждане на оператор на видимостта и анализ на неговото влияние върху правото и обратното разпространение на информацията в невронната мрежа.

Трета глава представя резултатите от два систематични компютърни експеримента, проведени в среда MATLAB върху колекция от 2748 изображения на военна техника в осем класа. В първия експеримент са оценени 19 предварително обучени невронни мрежи по отношение на базовата им точност при разпознаване на военни обекти. Вторият експеримент изследва ефективността на техниката за трансфер на знания и влиянието на аугментацията с частично закриване върху точността на разпознаване, като постигнатите резултати мотивират необходимостта от специализирана архитектура.

Четвърта глава съдържа основния научен принос на дисертацията — проектирането, систематичното изследване и оптимизацията на специализираната CNN-6 за разпознаване на частично видими военни обекти. Представени са резултатите от три експериментални етапа с нарастваща сложност на обучаващите данни, изборът и детайлното описание на оптималната шест-блокова архитектура, нейното сравнение с известни модели, разширението към задача за локализация чрез алгоритъм „плъзгащ прозорец“ и разработеният каскаден модел за семантично филтриране и разпознаване в смесени сцени с цивилни и военни обекти.

В **Приложение 1** в табличен вид и фигури са представени допълнителни данни от систематичният анализ на резултатите от работата на проектираните многослойните CNN.

В **Приложение 2** са приведени части от сорс-кодовете на програмите с помощта на които се осъществява моделирането, обучението и тестването на многослойните CNN.

2. КРАТКО СЪДЪРЖАНИЕ НА ДИСЕРТАЦИОННИЯТ ТРУД.

ГЛАВА 1. АНАЛИЗ НА СЪВРЕМЕННОТО СЪСТОЯНИЕ И ПРИЛОЖЕНИЕТО НА МЕТОДИТЕ НА ИЗКУСТВЕНИЯ ИНТЕЛЕКТ ЗА ОБРАБОТКА НА ОПТИЧНИ ИЗОБРАЖЕНИЯ.

Литературният обзор е изготвен с цел систематизиране и анализ на съществуващите научни изследвания, свързани с откриването и разпознаването на частично видими обекти с военно предназначение в оптични изображения. Обхваща 68 източника от последните петнадесет години с приоритет върху публикации след 2020 г., набрани от IEEE Xplore, SpringerLink, Elsevier Science Direct и Google Scholar.

1.1. Подходи за автоматично откриване и разпознаване на обекти в оптични изображения.

Автоматичното откриване и разпознаване на обекти е фундаментална задача в компютърното зрение с приложения в сигурността, отбраната, медицинската диагностика и автономните системи. Класическите методи, базирани на HOG, LBP, SIFT и SURF, показват съществени ограничения при сложни сцени — чувствителност към промени в осветеността, мащаба и оклузиите. С развитието на изчислителните ресурси методите, използващи многослойни невронни мрежи с дълбоко структурирано обучение се налагат като водещ подход, тъй като автоматично извличат йерархични характеристики без ръчно зададени признаци.

1.2. Дълбоки невронни мрежи за откриване и класификация на обекти.

CNN се утвърждават като основен инструмент за анализ на изображения - архитектурите еволюират от класификация на единични обекти към детекция на множество обекти с различни мащаби. Въпреки високата точност, CNN изискват значителни изчислителни ресурси и демонстрират понижена ефективност при сцени с оклузии и ограничени обучаващи данни.

1.3. Съвременни архитектури за откриване на обекти.

Съвременните подходи за откриване на обекти се делят на двустъпкови (Faster R-CNN) — с висока точност, но по-голяма изчислителна сложност — и едностъпкови (YOLO, SSD) - с ниска латентност и работа в реално време, но с по-ниска точност при малки или частично видими обекти. Повечето архитектури използват предварително обучени базови мрежи (VGG, ResNet, MobileNet), проектирани с общо предназначение и неоптимизирани за задачи с ограничена видимост.

1.4. Разпознаване на частично видими обекти и влияние на оклузиите.

Частичната видимост (оклузия) е едно от най-сериозните предизвикателства резултат от припокриване между обекти, препятствия или неблагоприятни ъгли на наблюдение. При оклузии стандартните CNN не могат да открият ключови структурни признаци, фокусират се върху вторични характеристики и генерират грешни решения. Предложените подходи - многомащабни представяния, механизми за внимание и контекстна информация частично решават проблема, но водят до увеличена сложност. Недостатъчното представяне на изображения на обекти с оклузии в обучаващите набори допълнително ограничава обобщаващата способност на моделите.

1.5. Трансфер на знания при задачи за откриване и разпознаване на обекти.

Обучението чрез трансфер на знания (Transfer Learning) е ключова техника при ограничени специфични данни - модели, предварително обучени върху големи бази от данни на изображения като ImageNet или COCO, се адаптират към целевата задача чрез допълнително обучение на част от слоевете на мрежата. При разпознаване на военни обекти разликата в изображенията спрямо общодостъпните набори е съществена, което може да доведе до отрицателен трансфер (Negative Transfer) - загуба на знания. Стратегиите за адаптация като замразяване на слоеве, постепенно допълнително обучение и диференцирани скорости на обучение изискват внимателна оптимизация.

1.6. Ограничения на съществуващите методи за откриване и разпознаване на обекти.

Анализът разкрива четири основни ограничения: (1) висока изчислителна сложност, затрудняваща работа в реално време; (2) недостатъчна устойчивост при оклузии; (3) зависимост от големи анотирани набори с ограничен достъп в специфични домейни; (4) ограничена интерпретируемост на решенията, затрудняваща анализа на грешките. Повечето съвременни архитектури са проектирани с приоритет върху точността, без оптимизация по ресурсна ефективност.

1.7. Роля на данните и анотациите при откриване и разпознаване на частично видими обекти.

В широко използваните бази от данни на изображения преобладават ясно видими обекти, което създава дисбаланс спрямо реалните условия. Анотирането при частична видимост е трудоемко и субективно — различията в интерпретацията водят до несъответствия, влияещи негативно върху

обучението. Предложените решения (частични анотации, маскиране, допълнителни класове за степен на оклузия) нямат универсален характер.

1.8. Симулиране на частична видимост и техники за увеличаване на обучаващите данни.

При ограничена наличност на реални изображения с оклузии се прилагат техники за аугментация — случайно маскиране, синтетични препятствия и комбиниране на обекти от различни изображения. Изкуствено генерираните примери не винаги отразяват реалните условия, поради което оптималният резултат се постига при комбиниране на синтетични и реални данни.

1.9. Метрики и критерии за оценка при откриване и разпознаване на обекти.

Стандартните метрики (Precision, Recall, F-мярка, mAP) не отразяват напълно резултатите от работата на многослойните невронни мрежи при частична видимост — частично коректните открития могат да бъдат отчетени като грешни. Прилагането на адаптивни прагове за припокриване и допълнителни показатели за устойчивост спрямо оклузии, осигурява по-реалистична оценка на поведението на моделите.

1.10. Устойчивост (робастност) и генерализация на моделите при сложни визуални условия.

Устойчивостта на моделите се изразява чрез способността им да поддържат стабилна ефективност при промени в условията на наблюдение — вариации в осветеността, мащаба, ъгъла и наличието на шум, оклузия. Генерализацията представлява способността на модела да прилага научените представяния върху непознати данни и да разпознава обекта, дори когато ключови визуални елементи липсват или са деформирани, което поставя високи изисквания към архитектурата и стратегията за обучение. Анализът на литературата показва, че прекомерно сложните модели постигат висока точност върху обучаващите данни, но демонстрират слаба генерализация при реални условия, което подчертава необходимостта от подходи, ограничаващи сложността и насърчаващи извличането на по-устойчиви признаци.

1.11. Влияние на фона и контекста на сцената върху разпознаването.

Фонът и контекстът на сцената играят двойствена роля — в едни случаи подпомагат правилната класификация, но в други водят до неправилни решения при прекомерна зависимост от вторични визуални елементи. Това е особено проблематично при военни обекти, които са умишлено замаскирани или използват камуфлаж, сливащ се с естествената среда, при което моделът може да се фокусира върху характеристики на фона вместо върху характерни признаци на обекта.

1.12. Изводи.

1. Анализът на описаните съществуващи архитектури, методи и подходи за откриване и разпознаване на обекти в изображения показва, че въпреки значителния напредък в областта, проблемът с частичната видимост остава нерешен.
2. Класическите методи за разпознаване на изображения, базирани на ръчно дефинирани признаци, показват ограничена устойчивост при сложни реални сцени.
3. С развитието на изчислителните ресурси и наличието на големи обеми от данни, постепенно се налагат методи, основани на машинно обучение и на многослойни невронни мрежи с дълбоко структурирано обучение.
4. CNN се утвърждават като основен инструмент за откриване и класификация на обекти в изображения.
5. Съвременните широко използвани CNN демонстрират ограничена ефективност при разпознаване на частично видими обекти.
6. Липсата на достатъчно примери на обекти с оклузия в обучаващите набори намалява устойчивостта на моделите в реални условия.
7. Предварително обучените модели (с бази от данни на изображения ImageNet и COCO) са оптимизирани за пълно видими обекти, което ограничава тяхната ефективност, особено при тежки оклузии.
8. Въпреки значителния напредък, съществуващите невронни архитектури срещат трудности при разпознаване на обекти в условия на оклузия и непълна информация. Това налага разработването на специализирани методи и оптимизирани архитектури, които да отчитат особеностите на подобни сцени и да балансират точност, устойчивост и изчислителна ефективност.

ГЛАВА 2. ТЕОРЕТИЧНИ ОСНОВИ И МАТЕМАТИЧЕСКИ МОДЕЛИ НА НЕВРОННИ МРЕЖИ С ДЪЛБОКО СТРУКТУРИРАНО ОБУЧЕНИЕ.

2.1. Обща характеристика и съвременно състояние на изследванията в областта на компютърното зрение, изкуствените невронни мрежи, машинното обучение и дълбоко структурираното обучение.

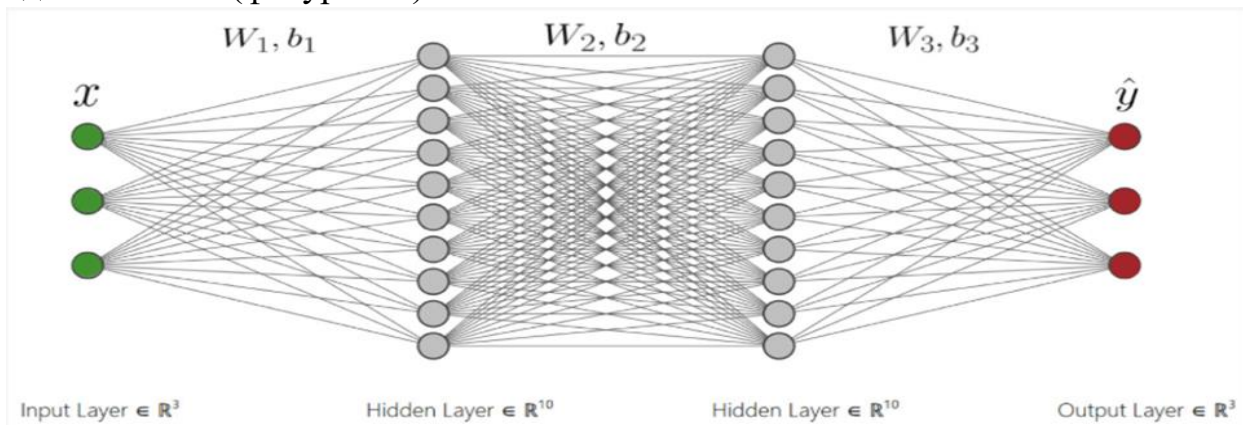
Компютърното зрение представлява интердисциплинарна област, която се занимава с разработването на методи за автоматично извличане, анализ и „разбиране“ на информацията от цифрови изображения или видеоклипове. Машинното обучение е подобласт на изкуствения интелект, в която системите имат способността автоматично да се учат и да се подобряват от опит без да бъдат изрично програмирани. Дълбоко структурираното обучение на

многослойни невронни мрежи, от своя страна, е специализиран клон на машинното обучение, който използва многослойни невронни мрежи за извличане на сложни абстракции от данните. Съвременната теория на компютърното зрение използва разнообразие от многослойни невронни архитектури с дълбоко структурирано обучение, всяка от които е проектирана за специфични задачи и предлага различни предимства. Актуалните изследвания са насочени към архитектурни иновации, оптимизация за устройства с ограничени ресурси, широко приложение на трансферното обучение и нарастващ фокус върху интерпретируемостта на моделите.

2.2. Характеристики на изкуствените невронни мрежи.

2.2.1. Устройство на многослойна невронна мрежа с право разпространение на информацията.

Многослойната невронна мрежа с право разпространение на информацията (feedforward neural network) се състои от входен слой, един или повече скрити слоя и изходен слой, като информацията се предава еднопосочно. (фигура 2.1)



Фиг.2.1. Многослойна невронна мрежа с напълно свързани слоеве и с право разпространение на информацията.

Източник: Feed-Forward Neural Networks for Failure Mechanics Problems. *Applied Sciences*, 11(14), 6483. Aldakheel, F., Satari, R., & Wriggers, P. (2021).

Изчисленията се описват чрез последователно прилагане на линейна трансформация и нелинейна функция на активация за всеки слой l :

$$(2.1) \quad z^{(l)} = W^{(l)} \cdot a^{(l-1)} + b^{(l)}, \quad l \in [1; L],$$

$$(2.2) \quad a^{(l)} = f^{(l)}(z^{(l)}),$$

$$(2.3) \quad a^{(0)} = x \text{ (входните данни),}$$

$$(2.4) \quad \hat{y} = a^{(L)} \text{ (предсказанието на мрежата).}$$

Съгласно теоремата за универсална апроксимация, мрежа с един скрит слой може да апроксимира произволна непрекъсната функция. „Дълбоките“ мрежи (с повече слоеве) могат да представят по-сложни функции в сравнение

с „плитките“ мрежи (1 или 2 слоя и повече неврони в слой). Композиционната структура на подобни многослойни мрежи предполага, че всеки следващ слой извлича все по-абстрактни представяния на входните данни.

2.2.2. Алгоритми и скорост на обучение.

Обучението на невронните мрежи се реализира чрез алгоритъм, адаптиращ теглата с цел минимизиране на грешката на изхода. Алгоритмите се класифицират като контролирано, неконтролирано и хибридно обучение. Скоростта на обучение (learning rate) е ключов параметър — много висока стойност води до нестабилност в резултатите от обучението, а твърде ниска — до бавна сходимост или засядане в локален минимум в процеса на обучение.

2.2.3. Алгоритъм за обучение чрез обратно разпространение на грешката в многослойни невронни мрежи.

Обратното разпространение на грешката е фундаментален алгоритъм за обучение, реализиран в два етапа - право разпространение на информацията за изчисляване на изхода и функцията на загубата:

$$(2.5) \quad z^{(l)} = W^{(l)}a^{(l-1)} + b^{(l)}, \quad l \in [1; L],$$

$$(2.6) \quad a^{(l)} = f^{(l)}(z^{(l)}), \quad l \in [1; L],$$

$$(2.7) \quad \hat{y} = a^{(L)},$$

$$(2.8) \quad L = -\sum_i y_i \log(\hat{y})_i.$$

Промяна на теглата на невроните чрез обратно разпространение на грешката с изчислените чрез верижното правило градиенти на параметрите:

$$(2.10) \quad \delta^{(L)} = \frac{\partial L}{\partial z^{(L)}} = (a^{(L)} - y) \odot f'^{(L)}(z^{(L)}),$$

$$(2.11) \quad \delta^{(l)} = (W^{(l+1)T} \delta^{(l+1)}) \odot f'^{(l)}(z^{(l)}),$$

$$(2.12) \quad \frac{\partial L}{\partial W^{(l)}} = \delta^{(l)} (a^{(l-1)})^T,$$

$$(2.13) \quad \frac{\partial L}{\partial b^{(l)}} = \delta^{(l)},$$

$$(2.14) \quad W^{(l)} := W^{(l)} - \eta \cdot \frac{\partial L}{\partial W^{(l)}},$$

$$(2.15) \quad b^{(l)} := b^{(l)} - \eta \cdot \frac{\partial L}{\partial b^{(l)}}.$$

Основни проблеми са „изчезващият“ и „експлодиращият“ градиент, решавани чрез функции на активация от тип ReLU, пакетна нормализация и прецизна инициализация. Оптимизациите включват мини-пакетно обучение, импулс (Momentum) и адаптивни алгоритми като Adam и RMSprop.

2.2.4. Функции на активиране, използвани при дълбоките невронни мрежи.

Изборът на активационна функция е критичен за ефективното обучение на многослойните невронни мрежи.

Функцията ReLU е стандартен избор за скритите слоеве на мрежата поради простота и устойчивост при проблема с „изчезващ“ градиент.

$$(2.19) \quad f(x) = \max(0, x).$$

Функциите Leaky ReLU и PReLU решават проблема с „умиращите“ неврони. Функцията GELU се използва в трансформерни архитектури.

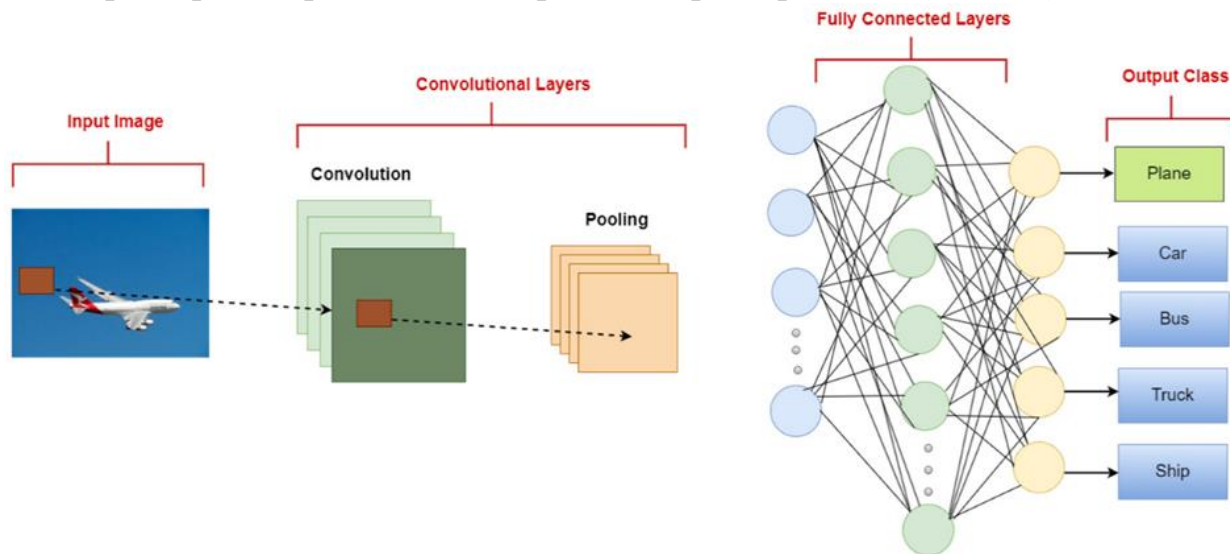
Функцията Softmax е стандарт за изходния слой при многокласова класификация.

$$(2.26) \quad f(z)_i = \frac{e^{(z_i)}}{\sum_{j=1}^n e^{(z_j)}}.$$

Изборът на активационна функция зависи от конкретната задача и архитектурата на мрежата.

2.3. Конволюционни невронни мрежи. Математически модел.

CNN са едни от най-популярните архитектури на невронни мрежи с приложение при многомерни входни данни, например изображения, видеоклипове, 3D модели. (фигура 2.2) При CNN се използват конволюционни слоеве, при които се извършва операцията конволюция с входните данни, чрез дефинирани филтри с цел открояване на пространствен контекст. По този начин CNN се обучават да класифицират изображения към съответни категории чрез откриване на абстрактни характеристики (features).



Фиг. 2.2. Стандартна CNN архитектура.

Източник: *Convolutional Neural Networks: A Comprehensive Evaluation and Benchmarking of Pooling Layer Variants. Symmetry, 16(11), 1516. Zafar, A. et al. (2024).*

2.3.1. Мрежови слоеве.

Предварителна обработка на входящите данни

Преди подаване към мрежата входните данни се нормализират чрез изваждане на средната стойност и деление със стандартното отклонение. Тази

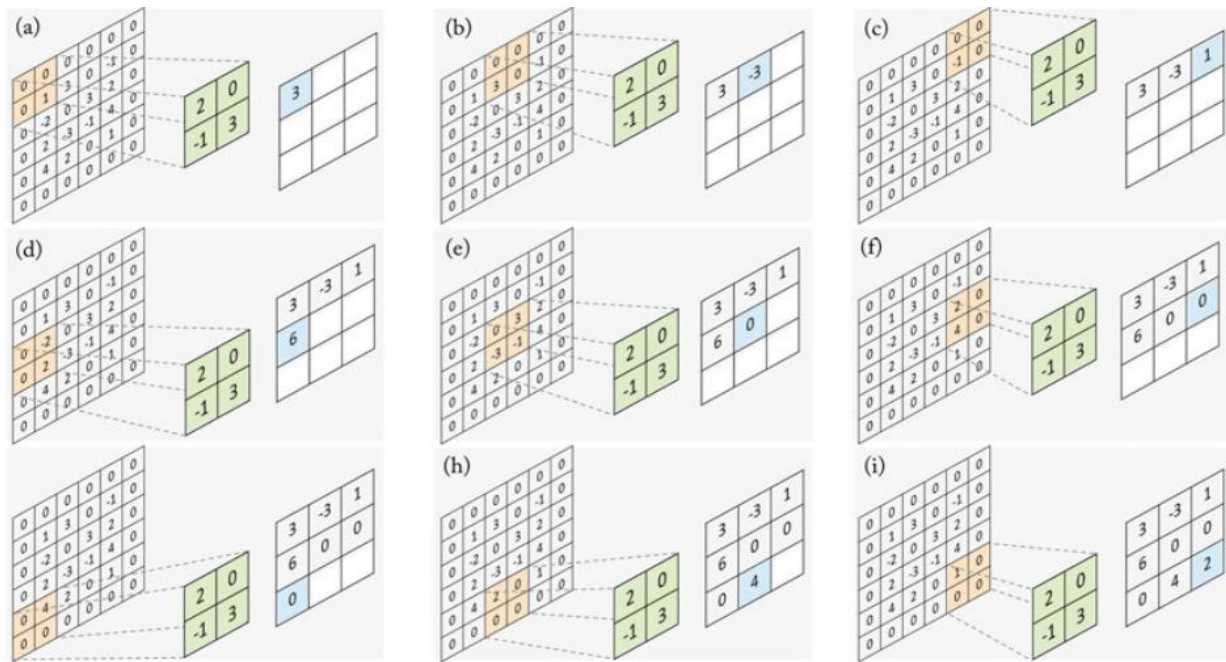
стъпка центрира данните спрямо нулата и осигурява равномерно мащабиране по всички размерности, което способства за стабилен процес на обучение при градиентни методи за оптимизация.

Конволюционни слоеве

Конволюционният слой се състои от набор от обучаеми филтри (ядра), представени като матрици, чиито стойности се адаптират в процеса на обучение. Всеки филтър се „плъзга“ по входните данни и генерира карта на характеристиките, отразяваща наличието на признак в различни части на изображението. Ключови параметри са размерът на филтъра f , стъпката s и допълването с нули p , като размерността на изходната карта се определя от израза (фигура 2.5):

$$(2.36) \quad h' = \left\lfloor \frac{h-f+s+p}{s} \right\rfloor,$$

$$(2.37) \quad w' = \left\lfloor \frac{w-f+s+p}{s} \right\rfloor.$$



Фиг. 2.5. Илюстрация на конволюционен слой с операция допълване с нули при $p=1$ и стъпка 2.

Използването на малки филтри (3×3 , 5×5) значително намалява броя на обучаемите параметри в сравнение с невронни мрежи с напълно свързани слоеве, като същевременно запазва способността за извличане на локални пространствени характеристики. Операцията допълване с нули (zero-padding) позволява запазване на пространствената размерност на изходната карта, което е важно при архитектури, изискващи точни предсказания по пиксели.

Рецептивно поле и разширена конволюция

Рецептивното поле определя пространствения обхват на всеки филтър спрямо входното изображение. При подреждане на n конволюционни слоя ефективното рецептивно поле нараства кумулативно. Стратегията за подреждане на много слоеве с малки филтри намалява броя на параметрите, но ограничава пространствения контекст, достъпен за всеки слой. За преодоляване на това ограничение се прилага разширената конволюция (Dilated Convolution) с коефициент d , която разширява ефективното рецептивно поле, без да увеличава броя на параметрите:

$$(2.42) \quad RF_{eff}^n = RF_{eff}^{n-1} + d(f - 1).$$

Това е особено важна техника при задачи за сегментация и разпознаване, изискващи едновременно локална прецизност и широк пространствен контекст — характеристики от съществено значение и при разпознаването на частично видими обекти.

Обединяващ слой (Pooling)

Слоят за обединяване редуцира пространствената резолюция на картите на характеристиките чрез прилагане на агрегираща функция (максимална или средна стойност) върху локални области с размер $f \times f$ и стъпка s . Операцията осигурява компактно представяне на характеристиките, което е инвариантно при умерени промени в мащаба, позицията и ориентацията на обекта в изображението. Чрез последователното намаляване на пространствената резолюция, слоевете за обединяване увеличават ефективния рецептивен обхват на следващите конволюционни слоеве и намаляват изчислителната сложност на мрежата.

Напълно свързани слоеве (Fully connected layers)

Напълно свързаните слоеве, разположени в края на архитектурата, интегрират извлечените пространствени характеристики в глобално представяне и реализират окончателната класификация. Всеки неврон в такъв слой е свързан с всички неврони на предходния слой, което позволява комбиниране на информация от различни части на изображението.

2.3.2. Функции за оценка на грешката в конволюционни невронни мрежи.

Функцията на загуба определя количествено разликата между прогнозирания и истинския изход на мрежата и се оптимизира по време на обучението чрез алгоритъма за обратно разпространение на грешката. Изборът на подходяща функция зависи от естеството на задачата. При многокласова класификация, основна функция е загубата на кръстосана ентропия,

използвана в настоящата дисертация и се прилага съвместно с функцията Softmax, която нормализира изходите към вероятностно разпределение:

$$(2.47) \quad L(p, y) = - \sum_{n=1}^N y_n \log(p_n),$$

$$(2.48) \quad p_n = \frac{e^{\hat{p}_n}}{\sum_{k=1}^N e^{\hat{p}_k}}.$$

Тази комбинация директно максимизира вероятността за правилна класификация и е стандарт при задачи с множество класове.

За регресионни задачи се използват Евклидовата загуба (средна квадратична грешка) l_2 и подобната l_1 грешка. Функцията Hinge Loss намира приложение при SVM-базирани класификатори, а контрастната загуба се използва при метрично обучение в сиамски мрежи за сравняване на двойки изображения.

За задачи по обработка и реконструкция на изображения е приложима мярката за структурно сходство SSIM, която оценява качеството на изображението не само на ниво пиксел, но и по отношение на структурни характеристики, яркост и контраст.

2.3.3. Обучение на конволюционни невронни мрежи (CNN Learning).

Инициализация на теглата

Правилната инициализация на теглата е предпоставка за стабилно и ефективно обучение. Нулевата инициализация е неприложима, поради проблема със симетрията - всички неврони в слоя биха се обновявали идентично. Гаусовата и равномерната случайна инициализация нарушават симетрията, но при многослойни невронни мрежи разпространяваните сигнали могат да нарастват или намаляват неконтролирано. Подобрена версия на инициализацията на Xavier е мащабираната ReLU инициализация с отчитане на мащаба и дава разпределението, от което произволно се избират началните тегла:

$$(2.60) \quad \text{Var}(w) = \frac{2}{n(f_{in})}.$$

Ортогоналната инициализация чрез SVD декомпозиция осигурява допълнителна стабилност при дълбоки мрежи.

Предварително обучение под наблюдение

Използването на трансферното обучение позволява адаптиране на предварително обучени модели към нови специфични задачи, като запазва извлечените общи визуални признаци от ниско ниво. При използване на готови архитектури (AlexNet, ResNet, DenseNet) предварително обучените тегла от ImageNet или COCO служат като начална точка, след което се

провежда фина настройка на последните слоеве с малка скорост на обучение. При персонализирани архитектури моделът се обучава върху мащабен набор от данни и след това се адаптира към целевата задача. Трансферното обучение е от особено значение при ограничени обучаващи данни, какъвто е случаят с разпознаването на военни обекти и е един от инструментите в експерименталната работа в настоящата дисертация.

Регуларизация

Многослойните невронни мрежи с голям брой параметри са склонни към преобучаване — постигат висока точност върху обучаващите данни, но се представят слабо върху непознати данни. За ограничаване на това явление се прилагат няколко подхода. Аугментацията на данните разширява обучаващия набор чрез геометрични трансформации (завъртания, мащабиране, огледално обръщане) и фотометрични вариации (промени в яркостта, контраста и цвета), като може да включва и синтетично генерирани данни.

Отпадане (Dropout)

Техниката на отпадането се прилага предимно за напълно свързани слоеве и деактивира неврони случайно чрез генерирана маска $m \in B^m$, където всеки елемент m_i е независимо избран от разпределението на Бернули с вероятност " p " да бъде "включен", т.е. се задейства неврон, което въвежда ефект на ансамбъл и намалява взаимната зависимост между невроните:

$$(2.61) \quad m_i \sim \text{Bernoulli}(p), \quad m_i \in m.$$

$$(2.62) \quad a^{(l)} = m \odot f(W^{(l)}a^{(l-1)} + b^{(l)}).$$

Нормализиране на пакети (Batch Normalization)

Пакетната нормализация (Batch Normalization) нормализира активациите на всеки слой към единично Гаусово разпределение, последвано от обучаемо преоразмеряване. Това стабилизира процеса на обучение, намалява чувствителността към избора на хиперпараметри, предотвратява проблема с „изчезващия“ градиент и значително ускорява сходимостта. Поради тези предимства пакетната нормализация се е наложила като стандартен компонент в съвременните CNN архитектури.

2.4. Други архитектури на невронни мрежи. Математически модели.

2.4.1. Рекурентни невронни мрежи (Recurrent Neural Network, RNN).

Рекурентните невронни мрежи са клас невронни мрежи, специално проектирани за обработка на последователни данни. RNN имат връзки, които формират цикли, позволявайки информацията да се запазва.

Математически модел:

$$(2.69) \quad h_t = f(W_{hh}h_{t-1} + W_{xt}x_t + b_h),$$

$$(2.70) \quad y_t = g(W_{hy}h_t + b_y).$$

Проблеми и решения:

При RNN се срещат проблемите с „изчезващия“ и “експлодиращия“ градиент при дълги последователности. Решенията включват: LSTM (Long Short-Term Memory), GRU (Gated Recurrent Unit), Attention механизми.

2.4.2. Автоенкодерите.

Автоенкодерите са невронни мрежи, които се учат да компресират данни в латентно представяне и след това да ги реконструират. Използват се за ненадзорано обучение, намаляване на размерността и извличане на характеристики.

2.4.3. Трансформери.

Трансформерите са архитектури, базирани на механизма на вниманието (attention), първоначално разработени за обработка на естествен език, но впоследствие намират приложение и в компютърното зрение.

Механизъм за самонаблюдение (Self-Attention):

За входни данни X , където $X \in R^{n \times d}$:

$$(2.74) \quad Q = XW^Q$$

$$(2.75) \quad K = XW^K$$

$$(2.76) \quad V = XW^V$$

$$(2.77) \quad \text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V.$$

W^Q , W^K , W^V са обучаеми параметри, които преобразуват X в заявки (Queries), ключове (Keys) и стойности (Values);

$$(2.78) \quad Q = XW^Q - \text{заявките } (n \times d_k);$$

$$(2.79) \quad K = XW^K - \text{ключовете } (n \times d_k);$$

$$(2.80) \quad V = XW^V - \text{стойностите } (n \times d_v),$$

като d_k е размерността на ключовете (2.79), а d_v - на стойностите.

2.4.4. Архитектури за детекция на обекти.

R-CNN, Fast R-CNN, Faster R-CNN, YOLO (You Only Look Once) и SSD (Single Shot Detector) са специализирани за локализация и класификация на множество обекти в изображения.

Примерен модел: YOLO - Разделя изображението на мрежа от координати с размери $s \times s$, за всяка клетка предсказва ограничителна рамка (Bounding Box) и класове.

$$(2.84) \quad y = [P_c C_1, P_c C_2, \dots, b_x, b_y, b_h, b_w],$$

P_c – достоверност за клас C , а (b_x, b_y, b_h, b_w) са координатите на ограничителната рамка.

2.4.5. Архитектури за семантична сегментация.

FCN (Fully Convolutional Networks), U-Net, SegNet и DeepLab са предназначени за пиксел-ниво класификация.

2.5. Математически модел за класификация на частично видими обекти чрез многослойна конволюционна мрежа.

Стандартният CNN модел приема, че входното изображение съдържа пълна информация за обекта. При реални военни сцени това допускане е нарушено — обектите са частично закрити от растителност, инфраструктура, камуфлаж или атмосферни явления.

2.5.1. Математичен модел на аугментирано изображение с оклузия.

Оптичното изображение се разглежда като двумерна векторна функция:

$$(2.86) \quad I(x, y): \Omega \subset R^2 \rightarrow R^3.$$

Всеки пиксел се представя като вектор от стойностите на цветовете канали и геометричните координати в равнината на изображението:

$$(2.87) \quad I(x, y) = [R(x, y), G(x, y), B(x, y)].$$

Оператор на видимостта – дефинира се чрез бинарна маска:

$$(2.88) \quad M(x, y): \Omega \rightarrow \{0, 1\},$$

$$(2.89) \quad M(x, y) = 1 \rightarrow \text{пикселът е наблюдаем},$$

$$(2.90) \quad M(x, y) = 0 \rightarrow \text{пикселът е закрит (оклузия)}.$$

Изображение с частична оклузия

$$(2.91) \quad I_M(x, y) = I(x, y) \odot M(x, y).$$

Задачата за разпознаване при наличие на оклузия е задача за класификация при наблюдение с непълна информация.

Коефициент на видимост

$$(2.92) \quad \eta = \frac{1}{|\Omega|} \int_{\Omega} M(x, y) dx dy$$

(в дискретен вид — процент не занулени пиксели).

CNN като вероятностен класификатор

CNN реализира функция:

$$(2.93) \quad F(I_M) = p(p_1, \dots, p_k),$$

(2.94) $p_k = (C_k | I_M)$ е вероятността изображението да принадлежи към C_k .

Класификацията се определя от:

$$(2.95) \quad \hat{C} = \arg \max_k (p_k).$$

Вероятност за правилна класификация

Точността на системата се разглежда като вероятност за правилно разпознаване $P_{correct}$ и зависи от видимата част на обекта:

$$(2.96) \quad P_{correct} = (\hat{C}(C_k | I_M)).$$

2.5.2. Право разпространение на информацията в невронни мрежи при изображения с оклузия.

$$(2.97) \quad I \in R^{H \times W \times C}$$

е изображението, с H -височина, W -ширина и C -цвет.

$$(2.98) \quad M \in \{0,1\}^{H \times W}$$

е маската на частична оклузията (O = скрит пиксел, I = видим пиксел).

$$(2.99) \quad I_M = I \odot M,$$

$$(2.100) \quad A^{(l)} = f(W^{(l)} * A^{(l-1)} + b^{(l)}),$$

$$(2.101) \quad A^{(0)} = I_M.$$

Тук приложението на маската се пропагира (предава през слоевете) непряко: оклузиите нулират входните пиксели, което променя изходните резултати във всички слоеве и води до различна финална вероятност за принадлежност към някой от класовете:

$$(2.102) \quad \hat{y} = \text{softmax}(A^{(L)}).$$

2.5.3 Алгоритъм за обратно разпространение на грешката за обучение на невронни мрежи в случаите на изображения с оклузия

Грешката в се изчислява като:

$$(2.103) \quad \mathcal{E} = -\sum_{c=1}^C y_c \log \hat{y}_c.$$

При обратното разпространение на грешката стойностите на градиентите се изчисляват спрямо маскираното чрез оклузия изображение:

$$(2.104) \quad \frac{\partial \mathcal{E}}{\partial W^{(l)}} = \frac{\partial \mathcal{E}}{\partial A^{(l)}} * A^{(l-1)}$$

$$(2.105) \quad \frac{\partial \mathcal{E}}{\partial A^{(l)}} = W^{(l)} * \frac{\partial \mathcal{E}}{\partial A^{(l)}}.$$

Тъй като

$$(2.106) \quad A^{(0)} = I \odot M,$$

градиентите също се маскират:

$$(2.107) \quad \frac{\partial \mathcal{E}}{\partial I_M} = \frac{\partial \mathcal{E}}{\partial A^{(0)}} * M.$$

Аугментацията на изображения чрез частични оклузии, които се използват в обучението на CNN, увеличава устойчивостта на мрежата към реални случаи на частично закрити обекти.

2.6. Изводи.

1. CNN са основните многослойни невронни мрежи използвани в компютърното зрение, благодарение на способността си да извличат пространствени характеристики чрез споделяне на тегла и локални рецептивни полета.
2. Разнообразието от слоеве (конволуционни, обединяване, напълно свързани, специализирани) позволява гъвкав дизайн на мрежи, адаптирани към специфични задачи.
3. Алгоритъмът за обратно разпространение на грешката остава фундаментален метод за обучение на многослойни невронни мрежи.
4. Изборът на активационна функция, правилната инициализация на теглата, и техниките за регуларизация като dropout, batch normalization и data

augmentation имат критично значение за успешното обучение на дълбоки мрежи.

5. Подходящият избор на функция за очаквана загуба зависи от естеството на задачата (класификация, регресия, метрично обучение) и има пряко влияние върху качеството на обучението.

6. Предварително обучените модели и финото настройване чрез трансферно обучение са подходи, които позволяват ефективна работа дори с ограничени данни.

7. Освен CNN, други архитектури като RNN, автоенкодери и трансформери разширяват възможностите на дълбоко структурираното обучение към различни домейни и задачи.

ГЛАВА 3. ОТКРИВАНЕ И РАЗПОЗНАВАНЕ НА ЧАСТИЧНО ВИДИМИ ОБЕКТИ С ВОЕННО ПРЕДНАЗНАЧЕНИЕ В ОПТИЧНИ ИЗОБРАЖЕНИЯ С ПОМОЩТА НА НЕВРОННИ МРЕЖИ.

Представен е систематичен анализ на работата на предварително обучени CNN, имплементирани в MATLAB, при разпознаване на военна техника в оптични изображения. Проведени са два взаимосвързани експеримента: оценка на базовата точност на оригиналните мрежи за класификация на изображения на военни обекти и приложение на трансфер на знания с цел подобряване на точността при класификация на частично видими военни обекти.

3.1. Материали и методи.

3.1.1. База от данни и подготовка на данните.

Използвана е колекция от 2748 изображения на военна техника, разпределени в осем класа: артилерия, БМП, БПЛА, бронирана машина, БТР, пехота, РСЗО и танк. Част от снимките са направени в реални бойни и тренировъчни условия, с присъствие на множество обекти и случаи на частично закриване. За управление на данните е използвана структурата ImageDatastore в MATLAB, осигуряваща ефективно зареждане и автоматично етикетирание без необходимост от едновременно зареждане на всички изображения в паметта.

3.1.2. Среда за числено експериментирание – MATLAB Deep Learning Toolbox.

В MATLAB с помощта на Deep Learning Toolbox™ се предоставят функции, приложения и блокове за проектиране, обучение, симулация и внедряване на многослойни невронни мрежи, в това число и CNN. Включените инструменти позволяват визуализация и анализ на работата на мрежите, проследяване на метрики за ефективност, прилагане на техники за оптимизация и внедряване на модели върху системи с ограничени ресурси.

3.1.3. Предварително обучени конволюционни невронни мрежи в MATLAB.

MATLAB Deep Learning Toolbox, предоставя 19 предварително обучени CNN архитектури — от класически (AlexNet, VGG-16/19, GoogLeNet) до съвременни (ResNet, DenseNet, EfficientNetB0, NasNet) и компактни (MobileNet-v2, ShuffleNet, SqueezeNet). Повечето са обучени с базата от данни на изображения ImageNet (1,2 млн. изображения, 1000 класа), единствено DarkNet-53 е предварително обучена с набора от данни на изображения COCO. Броят на параметрите на CNN варира от 1 до 143 милиона.

3.2. Първи експеримент — оценка на точността на предварително обучени невронни мрежи.

3.2.1. Методика на експеримента.

В рамките на първия експеримент е анализирана средната точност, минималната, максималната и средната достоверност на всичките 19 достъпни в MATLAB предварително тренирани мрежи, върху база от данни, съдържаща 2748 снимки на военна техника. Такава систематична оценка позволява да се изберат най-подходящи модели за конкретната задача и да се оптимизира тяхното използване в реални военни приложения.

3.2.2. Резултати от класификацията.

Всички 19 мрежи са тествани върху пълната база от данни на изображения без допълнително обучение. Средната достоверност варира между 54,56 % (EfficientNetB0) и 73,64 % (DenseNet201).

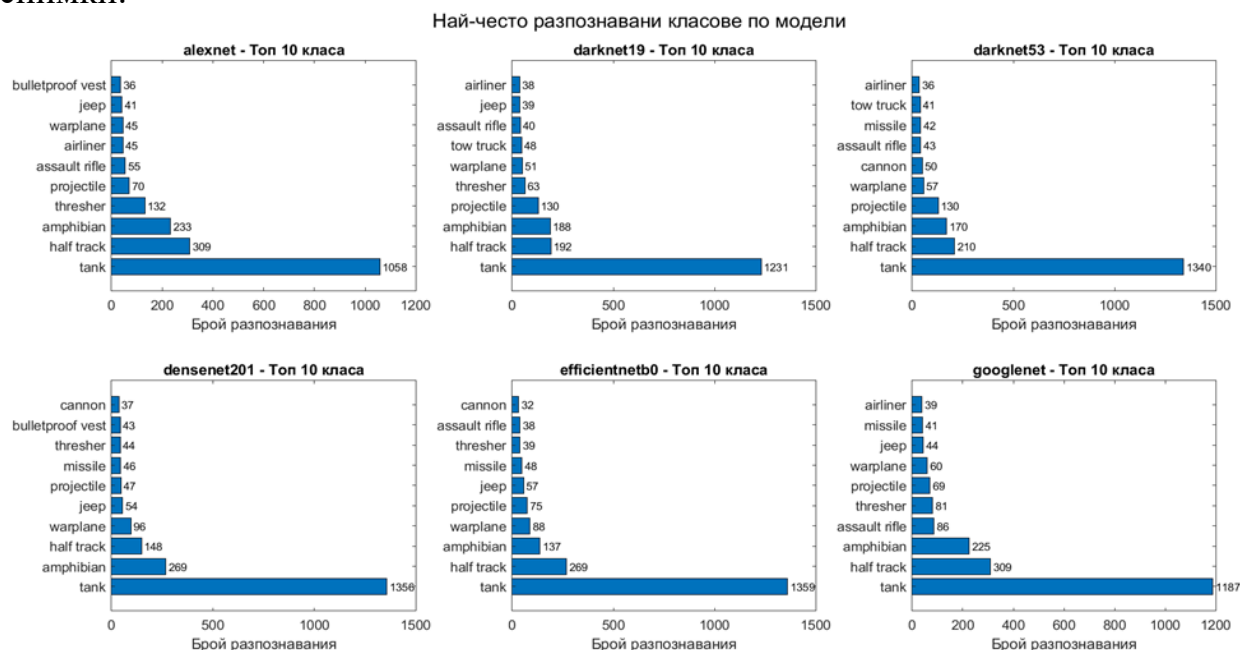
Таблица 3.3. Сравнителен анализ на резултатите от експеримента с оригинални предварително обучени мрежи.

Модел	Брой на изображения	Средна достоверност	Стандартно отклонение	Минимална достоверност	Максимална достоверност	Брой разпознати класове
ALEXNET	2748	62.71%	26.47%	4.23%	100%	192
VGG16	2748	57.82%	27.03%	4.82%	99.94%	156
VGG19	2748	59.27%	27.62%	3.15%	99.99%	143
GoogLeNet	2748	68.47%	25.37%	8.11%	100.00%	138
InceptionV3	2748	61.82%	25.75%	3.67%	99.54%	120
ResNet18	2748	62.10%	27.71%	4.84%	99.99%	154
ResNet50	2748	66.70%	27.56%	4.42%	100.00%	156
ResNet101	2748	70.90%	26.62%	7.30%	100.00%	136
DenseNet201	2748	73.64%	24.33%	7.98%	99.99%	121
MobileNetV2	2748	55.45%	27.70%	3.05%	99.93%	162
NASNetLarge	2748	70.37%	21.36%	8.79%	95.32%	105
NASNetMobile	2748	63.03%	25.09%	4.46%	99.17%	126
ShuffleNet	2748	58.52%	27.47%	4.55%	99.97%	175
SqueezeNet	2748	62.20%	26.22%	6.61%	99.99%	198
Inception ResNetV2	2748	68.61%	23.36%	10.69%	98.30%	103

Xception	2748	56.56%	27.42%	3.10%	99.98%	116
DarkNet19	2748	58.18%	29.43%	3.19%	99.97%	159
DarkNet53	2748	70.25%	26.39%	6.57%	100.00%	128
EfficientNetB0	2748	54.56%	26.01%	4.62%	98.54%	122

3.2.3. Анализ на най-често разпознаваните класове.

Всички разпознати класове са военни обекти или машини, които могат да се асоциират с военни обекти, или големи селскостопански машини, които при снимки с лошо качество могат да бъдат объркани с военни обекти (фигура 3.1). DarkNet53 е единствената мрежа, обучавана върху колекцията от снимки COCO. Статистическите ѝ параметри от експеримента са съпоставими с параметрите на другите 18 мрежи. При добре структурирано обучение сравними резултати могат да бъдат постигнати и с различни колекции от снимки.



Фиг. 3.1. Десет най-често разпознавани класа за шест избрани модела с различна архитектура.

3.2.4. Анализ на резултатите за точност.

В данните, представени в таблица 3.5, точността на мрежите е от един и същ порядък – не повече от 14.41 %.

Таблица 3.5. Точност за шест предварително тренирани мрежи с различна архитектура и данни за разпознат клас „танк“.

Мрежа	AlexNet	GoogLeNet	ResNet18	ResNet50	DarkNet19	EfficientNetB0	Средно
Точност	12.10%	14.26 %	13.47 %	14.02 %	14.06 %	14.41 %	13,72 %

Разпознат клас „танк“ /брой, %/	1058 38.55 %	1187 43.19 %	1227 44.65 %	1312 47.74 %	1231 44.79 %	1359 49.45 %	1229 44.72 %
---------------------------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------

Снимките на танкове, обозначени като такива в колекцията, са 415 и представляват само 15.10 % от всички. Анализът на данните показва, че самоходните артилерийски установки често биват класифицирани като „танк“, поради физически и визуални сходства, както и БТР, БМП и бронирани машини с въоръжение. Много от „научените“ от мрежите характеристики между военните обекти (машини) и някои тежки строителни и селскостопански машини също са сходни, което се вижда от общия процент разпознати класове, които могат да се асоциират с военни обекти (*фигура 3.1*).

3.3. Втори експеримент – трансфер на знания.

3.3.1. Методика на експеримент за трансфер на знания.

Приложен е трансфер на знания върху шест избрани архитектури (AlexNet, GoogLeNet, ResNet18, ResNet50, DarkNet19, EfficientNetB0) чрез замяна на изходните слоеве и допълнително обучение само на новите слоеве при „замразени“ останали тегла.

3.3.2. Прекласифициране и оценяване.

Оценяването включва сравнение на точността по класове преди и след допълнителното обучение на мрежите.

3.3.3. Подход за симулиране на частично закриване

За оценка на устойчивостта при оклузии оригиналната база е аугментирана чрез систематично закриване на изображенията с черни правоъгълници: 25 % закриване (по един квадрант) и 50 % закриване (по два квадранта), реализирано в четири ориентации. Подходът увеличава обема на базата от данни на изображения на военни обекти девет пъти и позволява контролирано тестване при различна степен на видимост.

3.3.4. Резултати от експеримента с трансфера на знания.

Трансферът на знания води до значително подобрение на точността при всички мрежи (таблица 3.6 и фигура 3.5). Най-висока точност върху оригиналните изображения постига EfficientNetB0 – 95.60 %, следвана от AlexNet (93.17 %) и ResNet50 (93.13 %). Изключение прави DarkNet19 с по-слабо подобрение от 40.10 %, а при GoogLeNet и DarkNet19 се наблюдава отрицателен трансфер за клас „танк“ (-20.7 % и -19.5 % съответно).

Таблица 3.6. Сравнителен анализ на обобщените резултати за точност при класифициране на изображения от оригиналната база от данни между предварително обучените (оригинални) и ретренираните чрез трансфер на знания невронни мрежи.

Мрежа	Оригинална, %	Ретренирана, %	Разлика, %
AlexNet	12.10	93.17	81.07

GoogLeNet	14.26	84.80	70.54
ResNet18	13.47	90.85	77.38
ResNet50	14.02	93.13	79.10
DarkNet19	14.06	54.16	40.10
EfficientNet-b0	14.41	95.60	81.19



Фиг. 3.5 Резултати от класифициране на военни обекти от ретренираната мрежа EfficientNetB0.

3.3.5. Резултати при частично видими обекти.

При класификация на аугментирани изображения (25 % и 50 % закриване) точността на ретренираните мрежи спада до диапазона 63 - 80 %, а EfficientNetB0 постига 80.31 %. Оригиналните предварително обучени мрежи се представят с 8 - 11 % точност върху аугментирани изображения (таблица 3.9).

Таблица 3.9. Сравнителен анализ на обобщените резултати за точност при класифициране на аугментирани с частично закриване изображения от ретренираните чрез трансфер на знания невронни мрежи.

Мрежа	Оригинална, %	Ретренирана, %	Разлика, %
AlexNet	7.98	68.54	60.57

GoogLeNet	9.56	63.04	53.49
ResNet18	9.50	69.18	59.68
ResNet50	10.27	72.48	62.22
DarkNet19	10.18	32.19	22.02
EfficientNet-b0	10.93	80.31	69.38

3.3.6. Резултати при трансфер на знания чрез обучение с аугментирани изображения.

Включването на аугментираните изображения в процеса на обучение води до съществено допълнително подобрение от ~ 24 %. Мрежата EfficientNetB0 (Net3), обучена върху аугментираната база, постига обща точност 99.25 % при разпознаване на клас „танк“ при 50 % закриване, в сравнение с 75.18 % на същата мрежа без аугментация в обучението (таблица 3.12). Резултатите са стабилни при всички варианти на закриване, включително при закрит долен или горен полукадър.

Таблица 3.12. Сравнителен анализ на обобщените резултати за точност по класове при класифициране на изображения от аугментираната база от данни на три различно обучени мрежи EfficientNetB0.

Име на клас	Различно обучени мрежи EfficientNet-b0		
	Net1 (Orig)	Net2 (No Aug)	Net3 (Aug)
artilleriya	0.00 %	69.70 %	97.67 %
bmp	0.00 %	76.78 %	95.40 %
bpla	0.00%	95.28 %	98.03 %
bronemashina	0.00 %	81.12 %	98.10 %
btr	0.00 %	75.05 %	96.45 %
pehota	0.00 %	91.96 %	96.76 %
rszo	0.00 %	81.94 %	98.71 %
tank	67.07 %	75.18 %	99.25 %

3.4. Изводи.

1. Въпреки значителните различия в архитектурите и в броя на параметрите (от 1 до 143 милиона), оригиналните предварително обучени невронни мрежи постигат сходни резултати по отношение на базовата точност (12 - 15 %) при разпознаване на военни обекти. Анализът на резултатите на най-често разпознаваните класове показва, че предварително обучените достъпни невронни мрежи имат тенденция да асоциират военните обекти със семантично близки категории като tank, half track, amphibian, warplane, предвид визуалното сходство на военната техника.

2. Техниката за трансфер на знания е ефективен метод за адаптиране на съществуващи модели към специфичната задача за разпознаване на военни

обекти, като подобрява точността с до *81.19 %* спрямо оригиналните мрежи. Най-добри резултати при разпознаване на оригинални изображения демонстрира невронна мрежа EfficientNetB0 с *95.60 %* точност след трансфер на знания, следвана от AlexNet (*93.17 %*) и ResNet50 (*93.13 %*) (таблица 3.6).

3. При разпознаване на частично видими обекти точността на предварително обучените невронни мрежи намалява и е в диапазона от *8 %* до *11 %*, а при ретренираните невронни мрежи с изображения без аугментация точността е в диапазона от *63 %* до *80 %* (таблица 3.9).

4. Обучението на невронните мрежи с аугментирани изображения с частично закрити обекти води до съществено повишаване на точността при класифициране на частично видими военни обекти – например при EfficientNetB0 се постига точност *99.25 %* при разпознаване на частично видими обекти от клас „танк“ (таблици 3.12 и 3.13).

5. Анализът на резултатите от проведените експерименти показва необходимост от създаване на специализирана конволюционна невронна архитектура, проектирана за разпознаване на военни обекти в оптични изображения при наличие на частично закриване, лоша видимост или присъствие на множество обекти в кадъра.

ГЛАВА 4. ПРОЕКТИРАНЕ НА КОНВОЛЮЦИОННА НЕВРОННА МРЕЖА ЗА ОТКРИВАНЕ И РАЗПОЗНАВАНЕ НА ЧАСТИЧНО ВИДИМИ ВОЕННИ ОБЕКТИ.

Четвърта глава представя проектирането, систематичното изследване и оптимизацията на специализирана конволюционна невронна архитектура за разпознаване на частично видими военни обекти. Основната цел е създаването на компактен модел с минимална архитектурна сложност, приложим в системи с ограничени изчислителни ресурси — бордови компютри на БПЛА, наземни наблюдателни станции и вградени платформи.

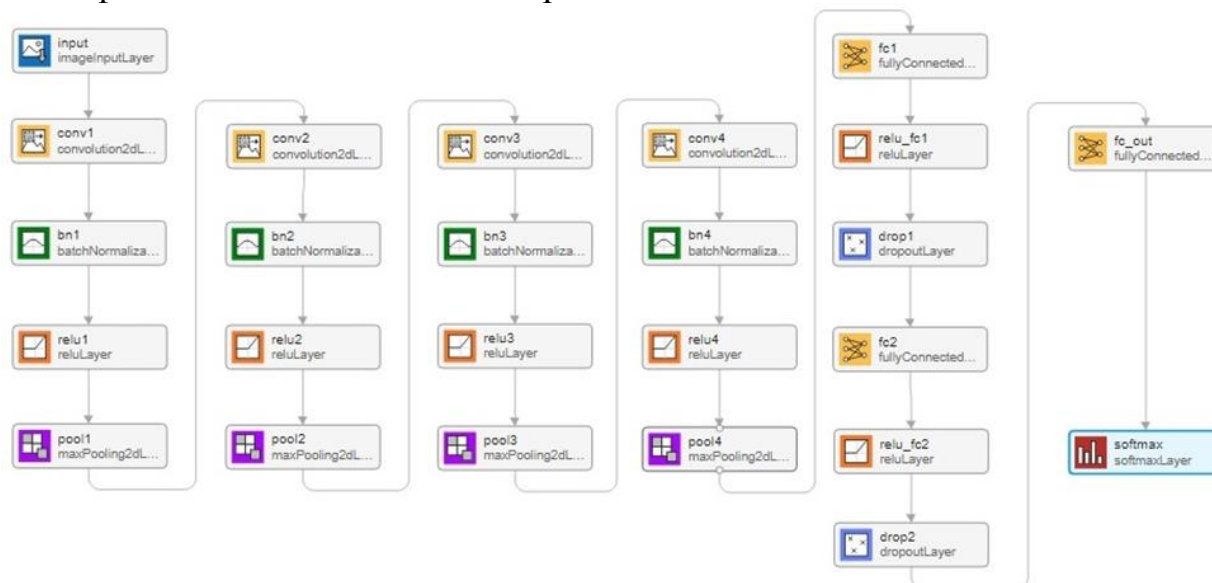
4.1. Проектиране на многослойни конволюционни невронни мрежи за задачата.

За осигуряване на съпоставимост всички проектирани архитектури работят с еднакъв входен формат на изображенията ($224 \times 224 \times 3$) и идентична изходна структура от осем класа. Изследвани са конфигурации с брой конволюционни блокове от 0 до 14, като систематичното изменение на дълбочината позволява изследване на ефекта на архитектурната сложност върху точността и устойчивостта при оклузия.

4.1.1. Базова конволюционна невронна мрежа с четири конволюционни блока.

Базовата архитектура се състои от четири последователни конволюционни блока, два напълно свързани блока и изходен слой. Всеки

конволюционен блок включва конволюционен слой (3×3 филтри), слой за пакетна нормализация, активационен слой (ReLU) и обединяващ слой с маска с размер (2×2) . Броят на филтрите нараства прогресивно: $8 \rightarrow 16 \rightarrow 32 \rightarrow 64$, а пространствената резолюция намалява: $224 \times 224 \times 3 \rightarrow 112 \times 112 \times 8 \rightarrow 56 \times 56 \times 16 \rightarrow 28 \times 28 \times 32 \rightarrow 14 \times 14 \times 64$. Напълно свързаните блокове съдържат 1024 и 512 неврона съответно, всеки с ReLU активация и dropout ($p = 0,4$). Изходният слой има 8 неврона, следван от слой със Softmax активация и класификационен слой. Общият брой слоеве е 26.



Фиг. 4.2. Четириблокова CNN (без класификационен слой).

Изборът на 4 конволюционни блока е теоретично обоснован: първият слой извлича нискостепенни признаци (контурни линии, текстури), вторият — фрагменти от форми и детайли на военната техника, третият — структурни представяния на компоненти (кула, оръдие, корпус), а четвъртият интегрира тази информация в абстрактни модели, способни да компенсират частичната видимост чрез характерни комбинации от видими признаци.

4.1.2. Модифицирани архитектури.

За систематично изследване на влиянието на архитектурната сложност, базовата мрежа е разширена в серия от модифицирани варианти с 3, 5, 6, 7, 8, 9, 10, 11, 12 и 14 конволюционни блока. Допълнително, за определени конфигурации са тествани варианти с различни размери на конволюционните филтри при фиксиран брой слоеве.

4.2. Методика на обучение и оценка на резултатите (4.2.1. — 4.2.3.).

Обучението и оценката са структурирани в три последователни експериментални етапа, всеки от които използва едни и същи архитектури при различни обеми и качество на обучаващите данни:

Етап 1 — обучение върху оригиналната колекция от 2748 изображения;

Етап 2 — обучение върху последователно аугментирани набори с геометрични и фотометрични трансформации ($\times 2$, $\times 4$, $\times 6$, $\times 8$ увеличение) без въвеждане на частична оклузия в обучаващите данни;

Етап 3 — обучение върху специално конструирана база с контролирана частична оклузия до 50 % (описана в Глава 3, раздел 3.3), симулираща реални сценарии на закриване.

4.3. Резултати от експериментите.

4.3.1. Общ подход за тестване и анализ на резултати.

Оценяването на всички архитектури се извършва върху единна тестова колекция с частично закрити изображения, включваща оригинали и варианти с 25 % и 50 % закриване в четири ориентации. Регистрираните метрики включват средна класификационна точност, разпределение на грешките по класове, деградация на точността спрямо оригинал и пространствена асиметрия при закриване.

4.3.2. Резултати от експериментите от етап 1.

При обучение само върху оригиналната неаугментирана колекция точността расте с броя конволюционни блокове до определен праг. CNN-12 постига най-висока точност при обучение (50.62 %), но с разлика от ~ 47 % между тренировъчната и валидационната крива — ясен признак на преобучаване (overfitting). По-плитките архитектури (≤ 7 блока) демонстрират ограничена способност за класификация при частична видимост, а по-дълбоките — нестабилни резултати. Средната точност върху аугментираната с оклузии колекция варира от 31 % (CNN-3) до 67 % (CNN-12).

4.3.3. Резултати от експериментите от етап 2.

Увеличаването на обема на обучаващите данни води до системно подобрение на точността при всички архитектури. При аугментация $\times 8$ CNN-12 постига 95.52 % точност при обучение и 74.43 % средна точност при частично закрити изображения. Допълнителното усложняване не носи пропорционален ефект - наблюдава се „плато“ при архитектури с повече от 7 – 8 блока. Максималната достигната средна точност при частична видимост в етап 2 остава по-малко от 75 % при всички архитектури, което потвърждава, че стандартната аугментация сама по себе си е недостатъчна.

4.3.4. Резултати от експериментите от етап 3.

Включването на контролирана оклузия в обучаващите данни води до качествена промяна в резултатите. Постигнатите стойности на точността при частично видими изображения нарастват значително при всички архитектури (таблица 4.8).

Наблюдава се ясно изразен максимум при архитектури с 6 – 8 конволюционни блока, след което допълнителното нарастване на дълбочината не носи систематично подобрение.

Таблица 4.8. Сравнителен анализ на точност от обучение и средна точност от класификация на изображения от база от данни с частично закриване в зависимост от броя конволюционни блока.

CNN-Брой конволюционни блока	3 блока	4 блока	5 блока	6 блока	7 блока	8 блока	9 блока	10 блока	11 блока	12 блока	14 блока
Точност при обучение с аугментирана колекция, %	78.14	86.58	85.99	92.08	93.56	94.17	93.82	93.01	92.22	95.42	92.44
Средна точност върху цялата колекция с частично закриване, %	82.41	90.92	91.87	96.97	97.87	97.82	97.71	97.05	96.87	98.21	96.88

4.4. Избор на оптимална архитектура.

4.4.1. Сравнение на анализ за избор на оптимална архитектура.

Изборът на оптимална архитектура се основава на три критерия: (1) средна класификационна точност по-голяма от 95 % при частично видими изображения; (2) постигане на тази точност при възможно най-малък брой конволюционни блокове; (3) стабилност — минимална разлика между точността при обучение и тази при тестване с изображения с оклузии.

Анализът на резултатите показва, че архитектурите с 3, 4 и 5 конволюционни блока не достигат прага от 95 %. При 6 конволюционни блока се наблюдава съществен скок — точността достига 96.97 % при стабилност 4.89 %. Архитектурите с 7 – 14 блока постигат сходни или малко по-високи стойности, но допълнителното подобрене не е пропорционално на нарасналата сложност.

Архитектурата CNN-6 с 6 конволюционни блока (34 невронни слоя) е определена като оптимален компромис между ефективност и архитектурна сложност — първата конфигурация, при която се достига висока и стабилна точност при класификация на частично видими военни обекти е представена на фигура 4.9.



Фиг. 4.9. Схема на реализираната CNN с шест конволюционни блока (без класификационен слой).

4.4.2. Детайлно описание и оптимизация на избраната архитектура CNN-6.

След избора на архитектурата е проведена двуетапна вътрешна оптимизация без промяна на броя слоеве.

Оптимизация на броя на конволюционните филтри: При фиксиран размер 3×3 са тествани различни конфигурации на броя и последователността на филтрите. Резултатите показват, че конфигурацията $32 \rightarrow 64 \rightarrow 128 \rightarrow 256 \rightarrow 512 \rightarrow 512$ постига най-висока средна точност — 98.92 % при частична оклузия, което представлява подобрене от 1.95 % спрямо изходната конфигурация $8 \rightarrow 16 \rightarrow 32 \rightarrow 64 \rightarrow 128 \rightarrow 256$.

Оптимизация на размера на първия конволюционен филтър: При фиксирана конфигурация на филтрите е тестван ефекта от промяната на размера на филтъра в първия слой (3×3 , 4×4 , 5×5 , 6×6 , 7×7). Резултатите показват, че размер 5×5 дава леко подобрене (98.97 %), интерпретирано като по-добър баланс между локално извличане на детайли и обхващане на по-широк пространствен контекст. При по-големи размери (6×6 , 7×7) точността слабо намалява, поради загуба на фини локални характеристики.

Финалната оптимизирана архитектура CNN-6 включва: входен слой ($224 \times 224 \times 3$), шест конволюционни блока с филтри 5×5 в първия и 3×3 в останалите, последователност на броя на филтрите $32 \rightarrow 64 \rightarrow 128 \rightarrow 256 \rightarrow 512 \rightarrow 512$, два напълно свързани блока (1024 и 512 неврона), Softmax и класификационен слой.

4.4.3. Резултати от детайлния анализ на оптимизираната конволюционна невронна мрежа с шест конволюционни блока.

Пълният статистически анализ по класове и варианти на оклузия е показан в таблица 4.12.

Таблица 4.12. Резултати за точност при класификация на изображения с оклузия по класове и варианти на оптимизираната CNN-6.

Варианти	Точност по класове и варианти, %							
	artilleriya	bmp	bpla	bronemashina	btr	pehota	rszo	tank
Original	100	99,40	99.26	99,12	99,33	99,51	100	100
Quad1_TopL	100	99,60	99.26	99,12	98,65	99,03	100	100
Quad2_TopR	100	99,00	99.26	99,41	100	99,51	100	100
Quad3_BottomL	100	99,40	99.26	99,41	98,65	99,03	100	100
Quad4_BottomR	100	98,80	99.26	99,41	99,33	99,51	100	100
Half1_Top	98,84	98,21	98.15	99,41	97,64	99,03	99,71	99,76
Half2_Bottom	97,67	95,02	95.94	97,35	94,61	97,57	99,13	94,70
Half3_Left	99,42	99,20	98.15	98,53	97,64	98,54	100	99,28
Half4_Right	98,84	98,01	99.26	98,53	98,32	99,51	100	100
Средна точност	99,42	98,52	98.65	98,92	98,24	99,03	99,87	99,30
Минимална	97,67	95,02	95.94	97,35	94,61	97,57	99,13	94,70
Максимална	100	99,60	99.26	99,41	100	99,51	100	100
Ст. отклонение	0,82	1,42	1.12	0,69	1,57	0,64	0,09	1,74

Тестването върху допълнителна база с нови варианти на закриване (5 %, 10 %, 15 %, 20 %, 75 %) показва, че мрежата успешно генерализира към варианти, за които не е обучавана. При 5 % закриване точността е ~ 99.5 %, при 10 % — ~ 99 %, при 15% — 94–97 %, при 20 % централно закриване — 85.66 %, а при 75 % закриване — 83–93 % в зависимост от ориентацията на закриването. (фигура 4.15)

Тези резултати потвърждават, че точността е нелинейна функция на коефициента на видимост η и притежава и качествено измерение — зависи не само от процента на закриване, но и от това коя конкретна част от обекта е закрита.

Допълнително тестване с импулсен шум (salt & pepper) с плътност 0.0–1.0 (таблица 4.14) показва, че мрежата поддържа приемлива класификационна точност при плътност до 0.4 (83.51 %). (фигура 4.16)

При анализа на резултатите следва да се има предвид, че операторът на оклузия, въведен в математическия модел на изображения с оклузия в Глава 2, тук има случайни стойности а не само 0 и 1. При правото разпространение на информацията в невронната мрежа това ще доведе до случайни активации и

нарушаване на запаметените при обучението корелационни връзки. Това се отразява на класификационната точност на невронната мрежа.



Фиг. 4.15. Допълнителни варианти на закриване.

Таблица 4.14. Сравнителен анализ на резултатите за точност на оптимизираната шестблокова CNN при класификация на частично закрити чрез прилагане на импулсен шум изображения.

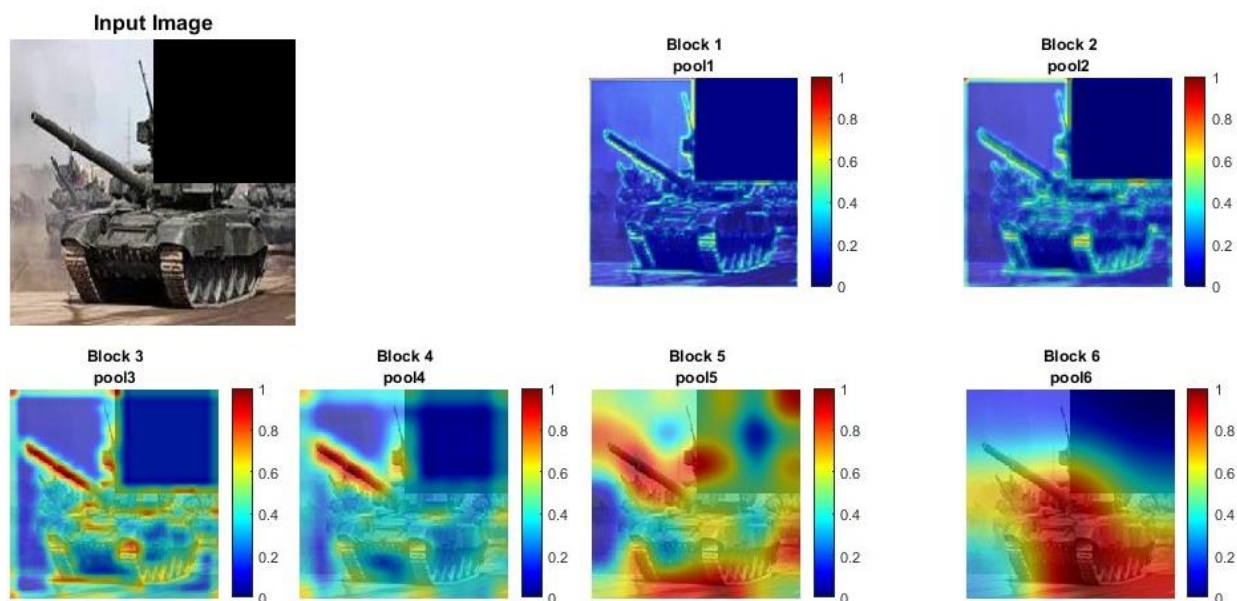
Стойност на параметъра за плътност на шум	Точност на класификация (%)	Най-труден за класифициране вариант, име и точност (%)
0,0	99,57	Quad1_TopLeft (99,56)
0,1	98,15	Half1_Top (94,58)
0,2	93,71	Half1_Top (76,79)
0,3	88,42	Half1_Top (63,16)
0,4	83,51	Half1_Top (53,65)
0,5	79,44	Half1_Top (47,60)
0,6	76,14	Half1_Top (43,01)

0,7	73,43	Half1_Top (40.22)
0,8	71,17	Half1_Top (37.86)
0,9	69,30	Half2_Bottom(34.96)
1,0	67,97	Half2_Bottom(32.29)



Фиг. 4.16. Варианти на закриване *Quad2_TopRight* за изображение на обект от клас „танк“ изпълнени чрез прилагане на функцията *salt&pepper* (импулсен шум) с различна плътност на шум (*noise density*).

На фигура 4.24 са представени входящото изображение и всички псевдоцветови карти на средната стойност по каналите на съответните последни слоеве в шестте блока, наложени върху входното изображение. Това се интерпретира като локални райони от изображението, които са важни за класификацията (осреднени характерни признаци, открити в работата на блока). Визуализацията показва как „вниманието“ на мрежата се концентрира върху различни детайли от входящото изображение и се извлича информация за комплексните структури в изображението.



Фиг. 4.24. Всички псевдоцветови карти на средната стойност по каналите (feature maps) на последните (pooling) слоеве в шестте блока наложени върху входното изображение.

4.5. Сравнение между ретренираната EfficientNet-B0 и CNN-6.

4.5.1. Архитектурна сложност

EfficientNet-B0 включва 237 слоя и 5,3 млн. обучаеми параметъра, докато CNN-6 има 34 слоя и 3,3 млн. параметъра. CNN-6 е проектирана без необходимост от предварително обучение върху мащабни набори от данни.

4.5.2. Сравнение при частична оклузия.

Сравнението е проведено на нископроизводителен процесор (Intel Pentium V960), мобилен процесор (Intel Core i3-5020U) и сървърна платформа (2×Intel Xeon Gold 6530). Резултатите (таблица 4.15) показват, че разликата в точността е под 1 %, докато времето за обработка на CNN-6 е 2–3 пъти по-кратко при ограничен хардуер. Това потвърждава, че специализирана и целево оптимизирана архитектура може да постигне резултати, съпоставими с наложени в практиката модели, при значително по-ниска изчислителна цена.

Таблица 4.15. Сравнителен анализ на параметри на ретренирана мрежа EfficientNet-B0 и CNN-6 получени при работа върху три различни хардуерни системи.

Параметър	EfficientNet-B0	CNN-6
Точност при оклузия	97.50%	96.97%
Размер на модела	14,97 MB	12,84 MB
Време на Pentium (s)	0,358	0,109
Време на Core i3 (s)	0,195	0,107
Време на Xeon (s)	0,302	0,157

4.6. Откриване на частично закрити военни обекти чрез имплементиране на алгоритъма „плъзгащ прозорец“ (sliding windows) и обучена шестблокова конволюционна мрежа за класификация.

За разширяване на CNN-6 към задача за локализация е имплементиран алгоритъм с „плъзгащ прозорец“ (sliding window). Изображението се сканира на части с различен мащаб и фиксирана стъпка; всеки регион се преоразмерява и подава за класификация. Прозорците с достоверност над предварително зададен праг формират двумерна карта на откриванията. Чрез запълване и обединяване на свързани региони се извлича ограничителна рамка (bounding box) около обекта. Допълнително за прецизиране на локализацията е използвана градиентно-базирана активационна карта (Grad-CAM). Подходът не изисква допълнително обучение с анотирани ограничителни рамки и е приложим при ограничени обучаващи данни.

4.7. Каскаден модел за семантично филтриране и разпознаване на частично видими военни обекти.

За реалистични сценарии с присъствие на цивилни и военни обекти едновременно е разработен каскаден модел, комбиниращ универсалната семантична мощ на EfficientNet-B0 с 1000 разпознаваеми класа и прецизната домейн-специфична класификация на CNN-6 за класове военна техника.

4.7.1. Теоретична постановка.

Моделът реализира йерархично филтриране в два основни етапа.

Етап 1А (семантичен „gatekeeper“): EfficientNet-B0 класифицира входното изображение. При разпознаване на известен цивилен ImageNet клас с достоверност над зададен праг, изображението се отхвърля като невоенно.

Етап 1Б (интервален анализ): От EfficientNet-B0 се извлича 1000-измерен вектор на характеристиките. *Основната хипотеза е, че това пространство съдържа имплицитна информация и за военни обекти, въпреки че те не са експлицитно дефинирани като класове.* Военните класове се моделират чрез статистически интервали в пространството на характеристиките. Само изображения с интервална оценка над определен праг τ се подават към Етап 2.

Етап 2: CNN-6 извършва детайлна класификация между осемте военни класа. При съответствие между CNN-6 и интервалния анализ се приема окончателна класификация, а при несъответствие се прилага претеглено решение. За потвърдените военни обекти се активира процедура за локализация чрез „плъзгащ“ прозорец.

4.7.2. Математически модел за интервален класификационен подход.

Изображение I се преобразува чрез предварително обучена невронна мрежа (feature extractor) във вектор в пространството на характеристиките:

$$(4.1) \quad x = f(I) \in R^d.$$

Модел на клас на военни обекти

За всеки клас $C_k, k = 1, 2, \dots, K$, се изчисляват статистики:

$$(4.2) \quad X_{k=(x^{(1)}, x^{(2)}, \dots, x^{(N_k)})}.$$

Вектор на математическо очакване

$$(4.3) \quad \mu_k = \frac{1}{N_k} \sum_{n=1}^{N_k} x^{(n)}.$$

Интервален модел

За всяка координата $i \in \{1, 2, \dots, d\}$ се определя интервал:

$$(4.4) \quad [l_{k,i}, u_{k,i}],$$

$$(4.5) \quad l_{k,i} = \min_n x_i^{(n)},$$

$$(4.6) \quad u_{k,i} = \max_n x_i^{(n)}.$$

Интервално съвпадение

За нов обект x се дефинира функция:

$$(4.7) \quad M_k(x) = \sum_{i=1}^d w_{k,i} \delta_{k,i}(x),$$

$$(4.8) \quad \delta_{k,i}(x) = \begin{cases} 1, & l_{k,i} \leq x_i \leq u_{k,i} \\ 0, & x_i < l_{k,i} \cup x_i > u_{k,i} \end{cases},$$

$$(4.9) \quad 0 \leq M_k(x) \leq 1, \quad w_{k,i} \geq 0, \quad \sum_{i=1}^d w_{k,i} = 1.$$

Разстояние до центъра на класа

$$(4.10) \quad D_k(x) = \sqrt{\sum_{i=1}^d w_{k,i} (x_i - \mu_{k,i})^2}.$$

Нормализирана оценка

$$(4.11) \quad S_k(x) = \frac{1}{1 + D_k(x)},$$

$$(4.12) \quad 0 < S_k < 1.$$

Комбинирана оценка

$$(4.13) \quad F_k(x) = \alpha M_k(x) + (1 - \alpha) S_k(x), \quad 0 \leq \alpha \leq 1.$$

Правило за класификация или отхвърляне

$$(4.14) \quad \hat{k} = \arg \max_k F_k(x), \quad F_{max}(x) = \max_k F_k(x).$$

При зададен праг $\tau \in \{0, 1\}$ крайното решение се формира като:

$$(4.15) \quad F_{max}(x) < \tau \Rightarrow x \neq \{1, 2, \dots, K\},$$

$$(4.16) \quad F_{max}(x) > \tau \Rightarrow x = \hat{k}.$$

4.7.3–4.7.7. Експериментална оценка и анализ на резултатите.

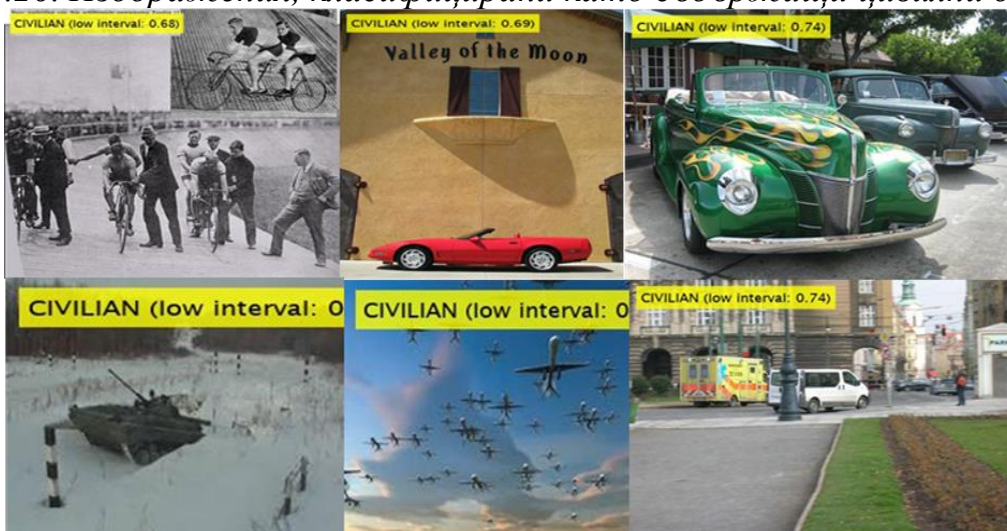
Каскадният модел е тестван върху 3047 изображения — 299 цивилни и 2748 военни. При прагове 0.80/0.75/0.90 резултатите показват:

- Етап 1А отхвърля като съдържащи цивилни обекти 150 изображения (4.9 %); (фигура 4.26)

- Етап 1Б отхвърля допълнително 111 изображения (3.6 %); (фигура 4.27)
- До CNN-6 достигат 2786 изображения (91.3 %);
- CNN-6 и интервалният анализ съвпадат при 58.2 % от изображенията;
- Финална класификация: 2699 военни (88.6 %), 261 цивилни (8.6 %), 87 неопределени (2.9 %) изображения;
- Ограничителна рамка е приложена за 99.9 % от класифицираните военни обекти.



Фиг. 4.26. Изображения, класифицирани като съдържащи цивилни обекти.



Фиг. 4.27. Изображения с максималната интервална оценка под предварително определен праг τ .

Класификационната точност на CNN-6 по отношение на военните обекти не се изменя при включването ѝ в каскадата, а времето за обработка на изображение на ниво система остава в допустими граници (1.87 s) дори при включена процедура за локализация.

Приносът на каскадния подход се изразява в аналитичното разширяване на функционалността на предварително обучена мрежа без промяна на научените тегла и без допълнително обучение — чрез детерминистичен механизъм, използващ вече наличните семантични знания в пространството на характеристиките. Това превръща големите предварително обучени архитектури в универсални семантични „сензори“, разширяеми към нови домейни с минимални изчислителни разходи.

4.8 Изводи.

1. Поетапният експериментален подход позволява систематично изследване на влиянието на архитектурната сложност (брой слоеве), обема на данните за обучение и наличието на частична оклузия върху точността и устойчивостта на CNN за разпознаване на военни обекти.
2. Увеличаването на дълбочината на архитектурата на невронната мрежа води до повишаване на класификационната точност само до определен праг, след който допълнителното усложняване не носи съществени ползи. Това обосновава избора на конволюционна архитектура с шест конволюционни блока (общо 34 слоя) като оптимален избор по отношение на постигната точност и изчислителна сложност.
3. Експериментите с умерена аугментация на обучаващите данни показват, че увеличаването на обема на обучаващият набор подобрява точността на модела.
4. Обучението чрез изображения с контролирана частична оклузия води до значително повишаване на точността и устойчивостта при реалистични изображения (таблици 4.8, 4.12 и 4.14).
5. Сравнителният анализ с ретренирани чрез трансферно обучение известни невронни мрежи показва, че шестблокова CNN постига съпоставима точност при значително по-малък брой на обучаемите параметри, по-ниска изчислителна сложност, по-малък размер на модела (обем на изпълнимия файл) и по-кратко време за обработка (таблица 4.15), което я прави подходяща за имплементиране в мобилни системи за работа в реално време.
6. Предложената CNN-6 архитектура може да бъде използвана и за задачи по откриване на военни обекти чрез имплементиране на алгоритъм „плъзгащ прозорец“ и анализ на локалните активации, без необходимост от допълнително обучение или използване на аотирани ограничителни рамки.
7. Разработеният каскаден модел за семантично филтриране комбинира предварително обучена универсална невронна мрежа, аналитичното

разширяване на семантичното пространство и специализирана конволюционна архитектура. Той представлява практически приложим подход за разграничаване между цивилни и военни обекти.

3. ЗАКЛЮЧЕНИЕ.

ОБЩИ ИЗВОДИ ОТ ДИСЕРТАЦИОННИЯТ ТРУД

1. В резултат на извършения аналитичен обзор на съществуващите методи за откриване и разпознаване на обекти се установи, че съвременните решения са ориентирани към цивилни приложения и не отчитат спецификата на военните обекти. Анализът показва съществени ограничения на тези методи при наличие на частично закриване, сложен фон и наличие на множество обекти в кадъра, което обоснова необходимостта от разработване и обучение на специализирана невронна мрежа.
2. Съществува необходимост от разработване на специализирана база от данни от оптични изображения на военна техника, включваща сцени от реални бойни и тренировъчни условия, частично закрити обекти и изображения с наличие на множество обекти. Базата от данни следва да осигурява необходимата представителност за обучение, тестване и обективна оценка на изследваните невронни архитектури.
3. На базата на проведеня анализ на съществуващите невронни архитектури, използвани в компютърното зрение, за постигане на поставената в дисертацията цел е обоснован избор на невронна архитектура от типа многослойна CNN.
4. Получените резултати от извършената систематична експериментална оценка на работата на предварително обучени известни CNN със структурираната специализирана база от данни служат за оценка на базовата класификационна способност и за идентифициране на техните ограничения по отношение на точността и устойчивостта при частично закриване.
5. Проектираната невронна архитектура с шест конволюционни блока се характеризира с подобрена точност при разпознаване на частично закрити военни обекти, както и по-добро съотношение между точност и изчислителна сложност в сравнение с изследваните претренирани известни невронни мрежи.
6. Разработеният каскаден модел за семантично филтриране, при който претренирана невронна мрежа се използва като първичен домейнов филтър, без да се извършва пълно преобучение върху разширен набор от класове, осигурява разграничаване между цивилни и военни обекти.
7. Проведеният сравнителен анализ между оптимизираната невронна мрежа и подобрените претренирани архитектури показва предимства на предложените решения по отношение на точност, устойчивост при частично

закриване и изчислителна ефективност. Получените експериментални резултати потвърждават, че специализираният архитектурен дизайн и каскадната организация на обработката са подходящи за задачи, свързани с разпознаване на военни обекти в реални условия.

В заключение може да се обобщи, че дисертационният труд предлага цялостен обоснован подход за откриване и разпознаване на частично видими военни обекти в оптични изображения. Представените архитектури, методи и алгоритми са приложими в системи за наблюдение и подпомагане на решенията в областта на отбраната, както и за бъдещо развитие на изследванията в тази насока.

4. ПРИНОСИ.

НАУЧНО-ПРИЛОЖНИ ПРИНОСИ В ДИСЕРТАЦИОННИЯ ТРУД

1. Осъществен е аналитичен обзор на съществуващите методи на изкуствения интелект за откриване и разпознаване на обекти в оптични изображения, включително такива на военна техника, както и при наличие на частично закриване на наблюдаваните в кадъра обекти.

2. Структурирана е специализирана база от данни с оптични изображения на военна техника, включваща сцени от реални бойни и тренировъчни условия, частично закрити обекти и изображения с множество цели в кадъра. Базата от данни е приспособена за обучение, тестване и оценка на работата на невронни мрежи при задачи, свързани с разпознаване на изображения с приложение във военната област (Глава 3).

3. Извършена е систематична експериментална оценка на съвременни предварително обучени известни невронни архитектури, върху специализираната база от данни, чрез което са установени техните възможности и ограничения при разпознаване на военни обекти, включително при условия на частично закриване и повишена сценична сложност (Глава 3).

4. Предложен е подход за обучение на специализирана архитектура на многослойна конволюционна невронна мрежа, предназначена за откриване и разпознаване на частично видими обекти с военно предназначение в оптични изображения. Избраният подход е с фокус върху устойчивост при частично закриване, сложен фон и наличие на множество обекти, което я отличава от стандартните решения, ориентирани към цивилни приложения (Глава 4).

5. Проектирана, реализирана и експериментално оптимизирана е CNN в развойната програмна среда на MATLAB, която се характеризира с по-добра устойчивост при частично закриване и по-добро съотношение между точност и изчислителна ефективност в сравнение с изследваните претренирани архитектури (таблици 4.(9-14) и фигури 4.(8-24)).

6. Предложен е каскаден модел за семантично филтриране, при който предварително тренирана невронна мрежа изпълнява ролята на първичен домейнов филтър, а специализираната невронна мрежа осъществява детайлната класификация на военните обекти. Моделът осигурява разграничаване на военни и цивилни обекти, без необходимост от пълно преобучение чрез разширен набор от класове (Глава 4, раздел 4.8).

7. Разработен е алгоритъм за разпознаване на допълнителни класове чрез вече налични признаци, извлечени от предварително тренирана невронна мрежа. Алгоритъмът използва статистически интервали и претеглено оценяване в пространството на признаците, без да се изменя първоначалната архитектура и без промяна на броя на изходните класове на мрежата (Глава 4 раздел 4.8.2, Приложение 2, програмен код 7).

8. Интегриран е алгоритъм за локализиране на военни обекти в изображения чрез имплементиране на „плъзгащ прозорец“, при което разработената невронна мрежа се използва като специализиран класификатор за откриване и класификация на частично закрити обекти (Приложение 2, програмен код 6 и 7).

9. Извършен е сравнителен анализ между предложената оптимизирана CNN и претренираните известни невронни архитектури, който експериментално потвърждава предимствата на предложената невронна архитектура и разработения каскаден подход по отношение на точност, устойчивост при наличие на частично закриване на наблюдаваните обекти и изчислителна ефективност (Глава 4, раздели 4.6 и 4.8).

ПРИЛОЖНИ ПРИНОСИ В ДИСЕРТАЦИОННИЯ ТРУД

1. Структурираната специализирана база от данни от оптични изображения на военни обекти, включваща сцени със сложен фон, частично закриване, различни мащаби и наличие на множество обекти в кадъра е аотирана и публикувана за свободен достъп в платформата Kaggle с цел да бъде използвана за тестване на невронни мрежи от други изследователи.

2. Разработен е софтуерен прототип в развойната програмна среда MATLAB, реализиращ предложените невронна архитектура и алгоритми, който може да бъде внедрен в системи за наблюдение и автоматизирано подпомагане на решенията в безпилотни летателни апарати за работа в реално време (Приложение 2).

3. Разработените програмни алгоритми са реализирани като модулни и разширяеми софтуерни компоненти, което позволява тяхното адаптиране към други задачи по обработка на изображения, разширяване към допълнителни класове обекти, използване на други типове сензорни данни или интеграция със съществуващи програмни и хардуерни платформи.

СПИСЪК С ПУБЛИКАЦИИТЕ:

[1] Манов, Л., & Кулев, Н. (2023), Откриване и разпознаване на частично видими обекти в оптични изображения с използване на техниките на изкуствения интелект - мини ревю от 1970 г. до 2000 г, *Proceedings of International Scientific Conference “Defense Technologies” DefTech 2023* (pp. 444–450), Faculty of Artillery, Air Defense and Communication and Information Systems, National Military University, ISSN 2367-7902.

[2] Манов, Л. Е. (2024), Съвременни технологии с изкуствен интелект за откриване на обекти с военно предназначение в оптични изображения, *Proceedings of International Scientific Conference “Defense Technologies” DefTech 2024* (pp. 675–682), Faculty of Artillery, Air Defense and Communication and Information Systems, National Military University, ISSN 2815-4282.

[3] Манов, Л. Е. (2025). Съвременни технологии с изкуствен интелект за откриване и разпознаване на частично видими обекти с военно предназначение в оптични изображения, *Proceedings of International Scientific Conference “Defense Technologies” DefTech 2025* (pp. 723–731), Faculty of Artillery, Air Defense and Communication and Information Systems, National Military University, ISSN 2815-4282.

[4] Манов, Л. Е. (2025), Съвременни технологии с изкуствен интелект за разпознаване на обекти с военно предназначение в оптични изображения: Компютърен експеримент с невронни мрежи в MATLAB, *Proceedings of International Scientific Conference “Defense Technologies” DefTech 2025* (pp. 714–722). Faculty of Artillery, Air Defense and Communication and Information Systems, National Military University, ISSN 2815-4282.

[5] Манов, Л. Е. (2025), Съвременни технологии с изкуствен интелект за откриване и разпознаване на частично видими обекти с военно предназначение в оптични изображения: Компютърен експеримент с трансфер на знания в MATLAB, *Proceedings of International Scientific Conference “Defense Technologies” DefTech 2025* (pp.732-741), Faculty of Artillery, Air Defense and Communication and Information Systems, National Military University, ISSN 2815-4282.



VASIL LEVSKI NATIONAL MILITARY UNIVERSITY
FACULTY OF "ARTILLERY, AIR DEFENCE AND COMMUNICATION AND
INFORMATION SYSTEMS"



9713 Shumen, 1 Karel Shkorpil Street
Tel: (054) 801 040; Tel./Fax: (054) 877 463
e-mail: nvu-sh@aadcf.nvu.bg

Department of "Computer Systems and Technologies"

Lyubomir Evgeniev Manov
Assistant Professor, Master of Physics

**DETECTION AND RECOGNITION OF PARTIALLY VISIBLE
MILITARY OBJECTS IN OPTICAL IMAGES USING ARTIFICIAL
INTELLIGENCE**

DISSERTATION ABSTRACT

submitted in fulfillment of the requirements for the award of the educational and
scientific degree "**Doctor**"

Doctoral Programme: "**Automated Systems for Information Processing and
Control**"

Field of Higher Education: **5. Technical Sciences**
Professional Field: 5.3. **Communication and Computer Engineering**

Academic Supervisor:
Col. Prof. Dr. Eng. Chavdar Nikolaev Minchev

Shumen, 2026.

The dissertation consists of 175 pages.
Number of tables: 29.
Number of figures: 48.
Number of appendices: 2.
Number of references: 136.
Number of publications related to the dissertation: 5.

The defence of the dissertation will be held on 11.05.2026 in the hall of ____.
Materials related to the defence are available to interested parties at ____, tel. ____.

The dissertation was discussed at a departmental council of the Department of Computer Systems and Technologies, Faculty of Artillery, Air Defence and Communication and Information Systems, Vasil Levski National Military University – Veliko Tarnovo, on 04.02.2026, and has been scheduled for defence before a scientific jury for the award of the educational and scientific degree of "Doctor".

The doctoral candidate is an Assistant Professor in the Department of Radar and Communication Systems, Faculty of Artillery, Air Defence and Communication and Information Systems, Vasil Levski National Military University – Veliko Tarnovo.

The main research activities of the dissertation were conducted in the MATLAB development environment.

Author: Assistant Professor Lyubomir Evgeniev Manov, Master of Physics
Title: "Detection and Recognition of Partially Visible Military Objects in Optical Images Using Artificial Intelligence"
Print run: 10 copies

1. GENERAL CHARACTERISTICS OF THE DISSERTATION

Rationale for the Dissertation

The motivation for this dissertation stems from the growing requirements placed on modern military surveillance systems and the inherent limitations of traditional methods for optical image processing. Under real-world conditions, objects of military and strategic importance are frequently partially occluded, camouflaged, or observed under adverse conditions. This results in low image quality and greatly hampers their reliable detection and recognition. There is a tangible need for methods and algorithms capable of providing higher accuracy and robustness in recognizing partially visible objects in complex environments.

Concurrently, advances in artificial intelligence and deep learning of multi-layer neural networks over the last decade have opened new possibilities for the automated processing of large volumes of visual information. Convolutional neural networks (CNNs) and related architectures have shown significant progress in solving object detection and recognition tasks in images; however, their application in the specific domain of military optical images with partial visibility remains insufficiently explored. This creates the preconditions for developing and investigating domain-adapted neural architectures, training strategies, and datasets tailored to the characteristics of this task.

An additional motivation is the need to increase the degree of automation and objectivity in decision-making at tactical and operational levels. By integrating artificial intelligence methods into surveillance and analysis systems, the goal is to reduce the workload on operators, limit subjective factors, and enhance the timeliness and reliability of the information provided.

Relevance of the Problem

In the context of growing threats from local and asymmetric armed conflicts, the need for continuous, reliable, and timely information about objects of military significance is of paramount importance. Modern military and intelligence systems generate enormous volumes of data from diverse sources – both military and civilian. A substantial portion of this information arrives in the form of optical images acquired at a distance from heterogeneous observation platforms – satellite systems, manned and unmanned aerial vehicles, and ground-based systems.

In practice, a significant proportion of these images is characterized by low quality and limited spatial resolution. They are obtained under adverse observation conditions and with partial visibility of the objects, due to occlusion, camouflage, or the presence of complex backgrounds. This significantly hampers traditional automated processing methods and demands the development of more robust approaches to detection and recognition.

To ensure the effective processing and interpretation of this information flow, it is necessary to deploy modern automated systems based on artificial intelligence

methods, and in particular deep learning of multi-layer neural networks. The ultimate goal is for the human decision-makers at local, tactical, operational, and strategic levels to have access to accurate, reliable, and timely information to support the planning and management of military operations.

Aim and Objectives of the Dissertation

The main aim of the dissertation is to design and optimise a deep-learning multi-layer neural network for the detection and recognition of partially visible military objects in optical images.

The objectives arising from the stated aim are:

1. To carry out an analysis of existing approaches for object detection and recognition in optical images using neural networks, with emphasis on defense applications.
2. To structure a specialized database of optical images of military equipment, including scenes with real combat and training conditions, with partial occlusion of objects and the presence of multiple objects per frame.
3. To perform a systematic evaluation of a set of modern pre-trained neural networks using the created database of optical images of military equipment.
4. To select a neural architecture for a multilayer neural network for object detection and recognition in optical images.
5. To design, implement, optimize, and evaluate a neural network for detection, recognition, and classification of partially occluded military objects in optical images.

Subject of the Dissertation

The subject of the dissertation is models and algorithms based on multilayer deep-learning neural networks for optical image recognition.

Means for Achieving the Stated Aim

- methods for digital image processing;
- theory and principles for building and operating artificial intelligence systems for image recognition;
- theory of computer vision;
- statistical methods for information processing;
- methods for software and simulation modelling using the MATLAB/Simulink development environment.

Scientific Novelty

The scientific novelty in the dissertation lies in the proposed approach for recognizing partially occluded military objects in optical images from a structured specialized database, using deep-learning neural architectures.

Practical Utility

The practical utility consists in the development of mathematical models and algorithms for detecting and recognizing partially visible military objects in optical

images, applicable in modern intelligence, surveillance, and targeting systems used by the armed forces and security structures, with the aim of increasing the reliability and speed of situational assessment at local, tactical, and operational levels.

Structure of the Dissertation

This dissertation develops and investigates approaches and architectures for automatic detection and recognition of partially visible military objects in optical images, based on multilayer deep-learning neural networks.

In the preface, the motives and relevance of the research are formulated, together with the subject of the scientific investigation, the aim and objectives of the dissertation, the used research methods, the scientific contributions, and the practical utility.

The dissertation contains a preface, four chapters, a conclusion, a list of scientific contributions, a list of publications, a list of cited literature, and appendices. The four chapters are logically interconnected and they follow the natural sequence of: analyzing the existing state of the problem, theoretically justifying the methods and models used, conducting systematic computer experiments, and formulating specific scientific contributions.

Chapter 1 presents a systematic literature review of the current state of methods for automatic object detection and recognition in optical images. Classical approaches based on hand-crafted features, modern deep neural architectures, and specific challenges related to partial visibility, occlusions, and limited training data in specialized domains are analyzed. The review covers 68 sources and serves as the basis for formulating the approach to solving the dissertation tasks.

Chapter 2 presents the theoretical foundations and mathematical models of multilayer deep-learning neural networks. The architecture and operating principles of CNNs, training and regularization algorithms, loss functions and main related architectures are examined. At the end of the chapter, a mathematical model for classifying partially visible objects is formalized by introducing a visibility operator and analyzing its influence on the forward and backward propagation of information in the neural network.

Chapter 3 presents the results of two systematic computer experiments conducted in MATLAB on a collection of 2,748 images of military equipment in eight classes. In the first experiment 19 pre-trained neural networks were evaluated for their baseline accuracy on military object recognition. The second experiment investigates the effectiveness of the transfer learning technique and the influence of occlusion augmentation on recognition accuracy, and the results motivate the need for a specialized architecture.

Chapter 4 contains the main scientific contribution of the dissertation — the design, systematic investigation, and optimization of the specialized CNN-6 for recognizing partially visible military objects. Results from three experimental stages

with increasing training data complexity are presented, along with the selection and detailed description of the optimal six-block architecture, its comparison with known models, extension to a localization task via a sliding window algorithm, and the developed cascade model for semantic filtering and recognition in mixed scenes with civilian and military objects.

Appendix 1 presents, in tabular and graphical form, supplementary data from the systematic analysis of results from the operation of the designed multilayer CNNs.

Appendix 2 provides parts of the source code of the programs used for modelling, training, and testing of the multilayer CNNs.

2. BRIEF CONTENT OF THE DISSERTATION

CHAPTER 1. ANALYSIS OF THE CURRENT STATE AND APPLICATION OF ARTIFICIAL INTELLIGENCE METHODS FOR OPTICAL IMAGE PROCESSING

The literature review was prepared to systematize and analyze existing scientific research related to the detection and recognition of partially visible military objects in optical images. It covers 68 sources from the last fifteen years, with priority given to publications after 2020, gathered from IEEE Xplore, SpringerLink, Elsevier Science Direct, and Google Scholar.

1.1. Approaches for Automatic Object Detection and Recognition in Optical Images

Automatic object detection and recognition is a fundamental task in computer vision, with applications in security, defense, medical diagnostics, and autonomous systems. Classical methods based on HOG, LBP, SIFT, and SURF show significant limitations in complex scenes — sensitivity to changes in illumination, scale, and occlusions. With the development of computational resources, methods using multilayer deep-learning neural networks have emerged as the leading approach, as they automatically extract hierarchical features without manually defined descriptors.

1.2. Deep Neural Networks for Object Detection and Classification

CNNs have established themselves as the primary tool for image analysis — architectures have evolved from single-object classification to detecting multiple objects at different scales. Despite high accuracy, CNNs require significant computational resources and demonstrate reduced effectiveness in scenes with occlusions and limited training data.

1.3. Modern Architectures for Object Detection

Modern object detection approaches are divided into two-stage (Faster R-CNN) — with high accuracy but greater computational complexity — and one-stage (YOLO, SSD) — with low latency and real-time operation, but lower accuracy for

small or partially visible objects. Most architectures use pre-trained backbone networks (VGG, ResNet, MobileNet) designed for general purposes and not optimized for limited-visibility tasks.

1.4. Recognition of Partially Visible Objects and the Effect of Occlusions

Partial visibility (occlusion) is one of the most serious challenges, resulting from overlap between objects, obstacles, or unfavorable observation angles. Under occlusions, standard CNNs cannot detect key structural features, focus on secondary characteristics, and generate incorrect decisions.

Proposed approaches — multi-scale representations, attention mechanisms, and contextual information — partially solve the problem but lead to increased complexity. Insufficient representation of occluded object images in training sets further limits model generalization capability.

1.5. Transfer Learning for Object Detection and Recognition Tasks

Transfer learning is a key technique when domain-specific data is limited. Pre-trained on large image datasets models such as ImageNet or COCO are adapted to the target task through additional training of some network layers. For military object recognition the domain gap relative to publicly available datasets is substantial, which may lead to negative transfer — loss of knowledge. Adaptation strategies such as layer freezing, gradual fine-tuning and differential learning rates require careful optimization.

1.6. Limitations of Existing Methods for Object Detection and Recognition

The analysis reveals four main limitations: (1) high computational complexity, hampering real-time operation; (2) insufficient robustness under occlusions; (3) dependence on large annotated datasets with limited access in specific domains; (4) limited decision interpretability, complicating error analysis. Most modern architectures are designed with accuracy as the primary objective, without optimization for resource efficiency.

1.7. Role of Data and Annotations in Detecting and Recognizing Partially Visible Objects

Widely used image databases are dominated by clearly visible objects, creating an imbalance relative to real-world conditions. Annotation under partial visibility is labor-intensive and subjective — differences in interpretation lead to inconsistencies that negatively affect training. Proposed solutions (partial annotations, masking, additional classes for degree of occlusion) lack universal applicability.

1.8. Simulating Partial Visibility and Data Augmentation Techniques

When real images with occlusions are limited, augmentation techniques are applied — random masking, synthetic obstacles, and combining objects from

different images. Artificially generated examples do not always reflect real-world conditions, so the optimal result is achieved by combining synthetic and real data.

1.9. Metrics and Evaluation Criteria for Object Detection and Recognition

Standard metrics (Precision, Recall, F-measure, mAP) do not fully reflect the performance of multilayer neural networks under partial visibility — partially correct detections may be counted as errors. Applying adaptive overlap thresholds and additional robustness indicators for occlusion provides a more realistic assessment of model behavior.

1.10. Robustness and Generalization of Models under Complex Visual Conditions

Model robustness is expressed through the ability to maintain stable performance under changes in observation conditions — variations in illumination, scale, angle, and the presence of noise and occlusion. Generalization is the ability of the model to apply learned representations to unseen data and recognize an object even when key visual elements are missing or distorted, placing high demands on architecture and training strategy. The literature analysis shows that overly complex models achieve high accuracy on training data but demonstrate poor generalization under real conditions, emphasizing the need for approaches that limit complexity and encourage extraction of more robust features.

1.11. Influence of Background and Scene Context on Recognition

The background and scene context play a dual role — in some cases supporting correct classification, but in others leading to incorrect decisions due to excessive dependence on secondary visual elements. This is especially problematic for military objects that are deliberately camouflaged or use camouflage that blends with the natural environment, where the model may focus on background characteristics instead of distinctive object features.

1.12. Conclusions

1. Analysis of the described existing architectures, methods, and approaches for object detection and recognition in images shows that, despite significant progress in the field, the problem of partial visibility remains unsolved.
2. Classical image recognition methods based on hand-crafted features show limited robustness in complex real-world scenes.
3. With the development of computational resources and the availability of large volumes of data, methods based on machine learning and deep learning of multilayer neural networks have progressively established themselves.
4. CNNs have established themselves as the primary tool for object detection and classification in images.
5. Widely used modern CNNs demonstrate limited effectiveness when recognizing partially visible objects.

6. The lack of sufficient examples of occluded objects in training sets reduces model robustness under real-world conditions.
7. Pre-trained models (with ImageNet and COCO image datasets) are optimized for fully visible objects, which limits their effectiveness, especially under severe occlusions.
8. Despite significant progress, existing neural architectures face difficulties in recognizing objects under occlusion and incomplete information. This necessitates the development of specialized methods and optimized architectures that account for the characteristics of such scenes and balance accuracy, robustness, and computational efficiency.

CHAPTER 2. THEORETICAL FOUNDATIONS AND MATHEMATICAL MODELS OF DEEP-LEARNING NEURAL NETWORKS

2.1. General Characteristics and Current State of Research in Computer Vision, Artificial Neural Networks, Machine Learning, and Deep Learning

Computer vision is an interdisciplinary field concerned with developing methods for the automatic extraction, analysis, and "understanding" of information from digital images or videos. Machine learning is a sub-field of artificial intelligence in which systems have the ability to automatically learn and improve from experience without being explicitly programmed. Deep learning of multilayer neural networks is a specialized branch of machine learning that uses multilayer neural networks to extract complex abstractions from data. Modern computer vision theory employs a variety of multilayer deep-learning neural architectures, each designed for specific tasks and offering different advantages. Current research is directed towards architectural innovations, optimization for resource-constrained devices, widespread application of transfer learning, and a growing focus on model interpretability.

2.2. Characteristics of Artificial Neural Networks

2.2.1. Structure of a Multilayer Feedforward Neural Network

A multilayer feedforward neural network consists of an input layer, one or more hidden layers, and an output layer, with information transmitted unidirectionally (Fig. 2.1).

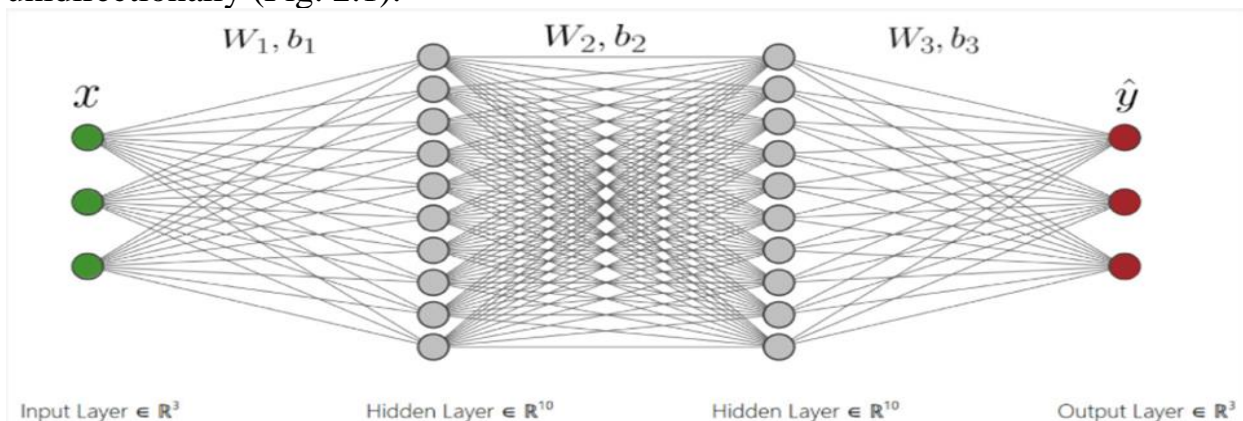


Fig. 2.1. Multilayer feedforward neural network with fully connected layers.

Source: Feed-Forward Neural Networks for Failure Mechanics Problems. *Applied Sciences*, 11(14), 6483.

Aldakheel, F., Satari, R., & Wriggers, P. (2021).

Calculations are described by the sequential application of a linear transformation and a nonlinear activation function for each layer l :

$$(2.1) \quad z^{(l)} = W^{(l)} \cdot a^{(l-1)} + b^{(l)}, \quad l \in [1; L],$$

$$(2.2) \quad a^{(l)} = f^{(l)}(z^{(l)}),$$

$$(2.3) \quad a^{(0)} = x \text{ (input data),}$$

$$(2.4) \quad \hat{y} = a^{(L)} \text{ (network prediction).}$$

According to the universal approximation theorem, a network with a single hidden layer can approximate any continuous function. Deep networks (with more layers) can represent more complex functions compared to "shallow" networks (1 or 2 layers with more neurons per layer). The compositional structure of such multilayer networks implies that each subsequent layer extracts increasingly abstract representations of the input data.

2.2.2. Training Algorithms and Learning Rate

Neural network training is performed using an algorithm that adapts the weights to minimize the output error. Algorithms are classified as supervised, unsupervised, and hybrid learning. The learning rate is a key parameter — an excessively high value leads to training instability, whereas a value that is too low results in slow convergence or entrapment in a local minimum during training.

2.2.3. Backpropagation Algorithm for Training Multilayer Neural Networks

Backpropagation is a fundamental training algorithm executed in two stages: forward propagation of information to calculate the output and loss function.

$$(2.5) \quad z^{(l)} = W^{(l)} a^{(l-1)} + b^{(l)}, \quad l \in [1; L],$$

$$(2.6) \quad a^{(l)} = f^{(l)}(z^{(l)}), \quad l \in [1; L],$$

$$(2.7) \quad \hat{y} = a^{(L)},$$

$$(2.8) \quad L = -\sum_i y_i \log(\hat{y})_i.$$

Weight adjustment via backpropagation, using gradients calculated through the chain rule:

$$(2.10) \quad \delta^{(L)} = \frac{\partial L}{\partial z^{(L)}} = (a^{(L)} - y) \odot f'^{(L)}(z^{(L)}),$$

$$(2.11) \quad \delta^{(l)} = (W^{(l+1)T} \delta^{(l+1)}) \odot f'^{(l)}(z^{(l)}),$$

$$(2.12) \quad \frac{\partial L}{\partial W^{(l)}} = \delta^{(l)} (a^{(l-1)})^T,$$

$$(2.13) \quad \frac{\partial L}{\partial b^{(l)}} = \delta^{(l)},$$

$$(2.14) \quad W^{(l)} := W^{(l)} - \eta \cdot \frac{\partial L}{\partial W^{(l)}},$$

$$(2.15) \quad b^{(l)} := b^{(l)} - \eta \cdot \frac{\partial L}{\partial b^{(l)}}.$$

Key challenges include the vanishing and exploding gradient problems, which are addressed through ReLU-type activation functions, batch normalization, and precise initialization techniques. Optimization strategies incorporate mini-batch learning, momentum, and adaptive algorithms such as Adam and RMSprop.

2.2.4. Activation Functions Used in Deep Neural Networks

The choice of activation function is critical for the effective training of multilayer neural networks. The ReLU function is the standard choice for hidden layers due to its simplicity and resistance to the "vanishing gradient" problem:

$$(2.19) \quad f(x) = \max(0, x).$$

Leaky ReLU and PReLU address the problem of "dying" neurons. The GELU function is used in transformer architectures. The Softmax function is the standard for the output layer in multi-class classification:

$$(2.26) \quad f(z)_i = \frac{e^{(z_i)}}{\sum_{j=1}^n e^{(z_j)}}.$$

The choice of activation function depends on the specific task and network architecture.

2.3. Convolutional Neural Networks. Mathematical Model

CNNs are among the most popular neural network architectures for multidimensional input data such as images, videos, and 3D models (Fig. 2.2). CNNs use convolutional layers where the convolution operation is performed on the input data through defined filters in order to highlight spatial context. In this way, CNNs are trained to classify images into corresponding categories by detecting abstract features.

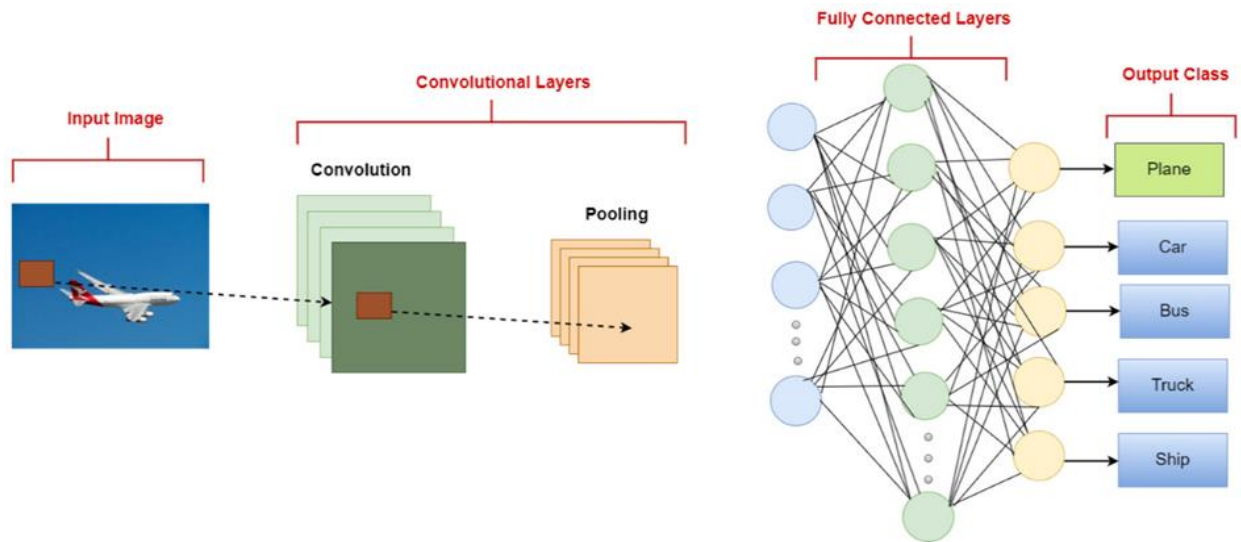


Fig. 2.2. Standard CNN architecture.

Source: Convolutional Neural Networks: A Comprehensive Evaluation and Benchmarking of Pooling Layer Variants. Symmetry, 16(11), 1516. Zafar, A. et al. (2024).

2.3.1. Network Layers

Pre-processing of Input Data

Before being fed to the network, the input data are normalized by subtracting the mean and dividing by the standard deviation. This step centres the data around zero and ensures uniform scaling across all dimensions, contributing to a stable training process for gradient-based optimization methods.

Convolutional Layers

A convolutional layer consists of a set of trainable filters (kernels) represented as matrices whose values adapt during training. Each filter "slides" over the input data and generates a feature map reflecting the presence of a feature in different parts of the image. Key parameters are the filter size f , stride s , and zero-padding p , and the output map dimension is determined by the expressions (Fig. 2.5):

$$(2.36) \quad h' = \left\lfloor \frac{h-f+s+p}{s} \right\rfloor,$$

$$(2.37) \quad w' = \left\lfloor \frac{w-f+s+p}{s} \right\rfloor.$$

The use of small filters (3×3 , 5×5) significantly reduces the number of trainable parameters compared to fully connected networks, while preserving the ability to extract local spatial features. The zero-padding operation allows the spatial dimension of the output map to be preserved, which is important in architectures that require pixel-accurate predictions.

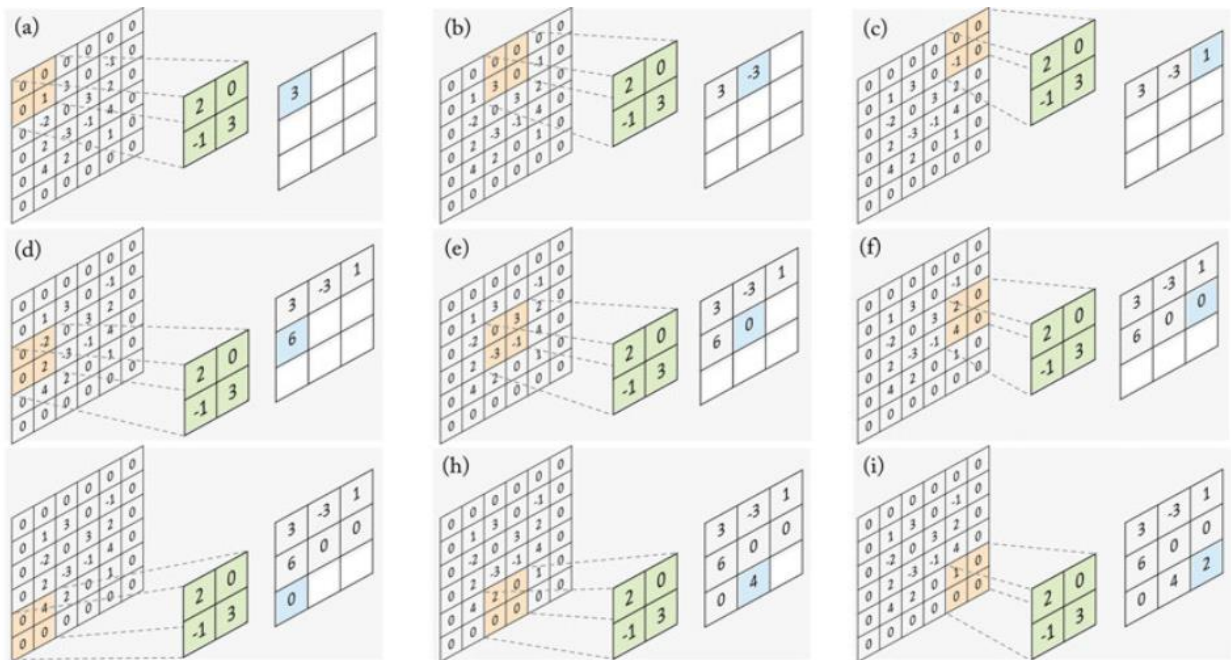


Fig. 2.5. Illustration of a convolutional layer with zero-padding at $p = 1$ and stride 2.

Receptive Field and Dilated Convolution

The receptive field determines the spatial scope of each filter relative to the input image. When stacking n convolutional layers, the effective receptive field grows cumulatively. The strategy of stacking many layers with small filters reduces the number of parameters but limits the spatial context available to each layer. To overcome this limitation, dilated convolution with coefficient d is applied, which expands the effective receptive field without increasing the number of parameters:

$$(2.42) \quad RF_{eff}^n = RF_{eff}^{n-1} + d(f - 1).$$

This is a particularly important technique for segmentation and recognition tasks requiring simultaneous local precision and wide spatial context — characteristics of significant relevance to the recognition of partially visible objects.

Pooling Layer

The pooling layer reduces the spatial resolution of the feature maps by applying an aggregation function (maximum or average) over local regions of size $f \times f$ with stride s . The operation provides a compact representation of features that is invariant to moderate changes in the object's scale, position, and orientation. By successively reducing spatial resolution, pooling layers increase the effective receptive scope of subsequent convolutional layers and reduce the computational complexity of the network.

Fully Connected Layers

Fully connected layers, located at the end of the architecture, integrate the extracted spatial features into a global representation and perform the final classification. Each neuron in such a layer is connected to all neurons of the previous layer, allowing information from different parts of the image to be combined.

2.3.2. Loss Functions in Convolutional Neural Networks

The loss function quantifies the difference between the predicted and true output of the network and is optimized during training through the backpropagation algorithm. The choice of an appropriate function depends on the nature of the task. For multi-class classification, the primary function is the cross-entropy loss, used in this dissertation, applied jointly with the Softmax function, which normalizes the outputs to a probability distribution:

$$(2.47) \quad L(p, y) = - \sum_{n=1}^N y_n \log(p_n),$$

$$(2.48) \quad p_n = \frac{e^{\hat{p}_n}}{\sum_{k=1}^N e^{\hat{p}_k}}.$$

This combination directly maximizes the probability of correct classification and is the standard for multi-class tasks.

For regression tasks, the Euclidean loss (mean squared error) l_2 and the similar l_1 error are used. Hinge Loss is applied in SVM-based classifiers, and contrastive loss is used in metric learning in Siamese networks for comparing image pairs. For image processing and reconstruction tasks, the Structural Similarity Index Measure (SSIM) is applicable.

2.3.3. Training of Convolutional Neural Networks (CNN Learning)

Weight Initialization

Proper weight initialization is a prerequisite for stable and effective training. Zero initialization is inapplicable due to the symmetry problem. Gaussian and uniform random initialization break symmetry, but for multilayer networks the propagated signals may grow or decrease uncontrollably. An improved version of Xavier initialization is the scaled ReLU initialization:

$$(2.60) \quad \text{Var}(w) = \frac{2}{n(f_{in})}.$$

Orthogonal initialization via SVD decomposition provides additional stability for deep networks.

Supervised Pre-training

Transfer learning allows pre-trained models to be adapted to new specific tasks, preserving extracted general low-level visual features. When using established architectures (AlexNet, ResNet, DenseNet), pre-trained weights from ImageNet or COCO serve as a starting point, after which fine-tuning of the last layers is conducted with a small learning rate. Transfer learning is particularly important with limited training data, as is the case with military object recognition, and is one of the tools in the experimental work in this dissertation.

Regularization

Multilayer neural networks with a large number of parameters are prone to overfitting — achieving high accuracy on training data but performing poorly on unseen data. To limit this, several approaches are applied. Data augmentation extends the training set through geometric transformations (rotations, scaling, flipping) and photometric variations (brightness, contrast, colour changes), and may include synthetically generated data.

Dropout

The dropout technique is applied primarily to fully connected layers and randomly deactivates neurons through a generated mask $m \in B^m$, where each element m_i is independently drawn from a Bernoulli distribution with probability p of being "active", introducing an ensemble effect and reducing co-dependency between neurons:

$$(2.61) \quad m_i \sim \text{Bernoulli}(p), \quad m_i \in m.$$

$$(2.62) \quad a^{(l)} = m \odot f(W^{(l)}a^{(l-1)} + b^l).$$

Batch Normalization

Batch normalization normalizes the activations of each layer to a unit Gaussian distribution, followed by a learnable rescaling. This stabilizes the training process, reduces sensitivity to the choice of hyperparameters, prevents the "vanishing gradient" problem, and significantly accelerates convergence. For these reasons, batch normalization has become a standard component in modern CNN architectures.

2.4. Other Neural Network Architectures. Mathematical Models

2.4.1. Recurrent Neural Networks (RNN)

Recurrent neural networks are a class of neural networks specifically designed for processing sequential data. RNNs have connections that form cycles, allowing information to be preserved. Mathematical model is:

$$(2.69) \quad h_t = f(W_{hh}h_{t-1} + W_{xt}x_t + b_h),$$

$$(2.70) \quad y_t = g(W_{hy}h_t + b_y).$$

RNNs encounter the problems of "vanishing" and "exploding" gradients for long sequences. Solutions include: LSTM (Long Short-Term Memory), GRU (Gated Recurrent Unit), and Attention mechanisms.

2.4.2. Autoencoders

Autoencoders are neural networks that learn to compress data into a latent representation and then reconstruct it. They are used for unsupervised learning, dimensionality reduction, and feature extraction.

2.4.3. Transformers

Transformers are architectures based on the attention mechanism, originally developed for natural language processing, but subsequently finding application in computer vision. Self-attention mechanism for input X , where $X \in R^{n \times d}$:

$$(2.74) \quad Q = XW^Q$$

$$(2.75) \quad K = XW^K$$

$$(2.76) \quad V = XW^V$$

$$(2.77) \quad \text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V.$$

W^Q, W^K, W^V are trainable parameters that transform X into queries, keys and values:

$$(2.78) \quad Q = XW^Q - \text{запросы } (n \times d_k);$$

$$(2.79) \quad K = XW^K - \text{ключевые } (n \times d_k);$$

$$(2.80) \quad V = XW^V - \text{значения } (n \times d_v),$$

where d_k is the dimension of the keys (2.79) and d_v is of the values (2.80).

2.4.4. Architectures for Object Detection

R-CNN, Fast R-CNN, Faster R-CNN, YOLO (You Only Look Once), and SSD (Single Shot Detector) are specialized for the localization and classification of

multiple objects in images. Example model: YOLO divides the image into an $s \times s$ coordinate grid; for each cell it predicts a bounding box and classes:

$$(2.84) \quad y = [P_c C_1, P_c C_2, \dots, b_x, b_y, b_h, b_w],$$

where P_c is confidence for class C , and (b_x, b_y, b_h, b_w) are the bounding box coordinates.

2.4.5. Architectures for Semantic Segmentation

FCN (Fully Convolutional Networks), U-Net, SegNet, and DeepLab are designed for pixel-level classification.

2.5. Mathematical Model for Classifying Partially Visible Objects via a Multilayer Convolutional Network

The standard CNN model assumes that the input image contains complete information about the object. In real military scenes this assumption is violated — objects are partially occluded by vegetation, infrastructure, camouflage, or atmospheric phenomena.

2.5.1. Mathematical Model of an Augmented Image with Occlusion

The optical image is presented as a two-dimensional vector function:

$$(2.86) \quad I(x, y): \Omega \subset R^2 \rightarrow R^3.$$

Each pixel is represented as a vector of color channel values and geometric coordinates in the image plane:

$$(2.87) \quad I(x, y) = [R(x, y), G(x, y), B(x, y)].$$

Visibility Operator - defined through a binary mask:

$$(2.88) \quad M(x, y): \Omega \rightarrow \{0, 1\},$$

$$(2.89) \quad M(x, y) = 1 \rightarrow \text{пикселът е наблюдаем},$$

$$(2.90) \quad M(x, y) = 0 \rightarrow \text{пикселът е закрит (оклужия)}.$$

Image with Partial Occlusion

$$(2.91) \quad I_M(x, y) = I(x, y) \odot M(x, y).$$

The recognition task in the presence of occlusion is a classification task under observation with incomplete information.

Visibility Coefficient

$$(2.92) \quad \eta = \frac{1}{|\Omega|} \int_{\Omega} M(x, y) dx dy$$

(in discrete form — the percentage of non-zeroed pixels).

CNN as a Probabilistic Classifier

The CNN implements a function:

$$(2.93) \quad F(I_M) = p(p_1, \dots, p_k),$$

$$(2.94) \quad p_k = (C_k | I_M) \text{ is the probability of the image belonging to}$$

C_k .

Classification is determined by:

$$(2.95) \quad \hat{C} = \arg \max_k (p_k).$$

Probability of Correct Classification

The accuracy of the system is viewed as the probability of correct recognition $P_{correct}$ and depends on the visible part of the object:

$$(2.96) \quad P_{correct} = (\hat{C}(C_k|I_M)).$$

2.5.2. Forward Propagation in Neural Networks for Images with Occlusion

(2.97) $I \in R^{H \times W \times C}$ — image with height H, width W, and C colour channels.

(2.98) $M \in \{0,1\}^{H \times W}$ — partial occlusion mask (0 = occluded pixel, 1 = visible pixel).

$$(2.99) \quad I_M = I \odot M,$$

$$(2.100) \quad A^{(l)} = f(W^{(l)} * A^{(l-1)} + b^{(l)}),$$

$$(2.101) \quad A^{(0)} = I_M.$$

Here the application of the mask propagates indirectly through the layers: occlusions zero out the input pixels, changing the output results in all layers and leading to a different final class membership probability:

$$(2.102) \quad \hat{y} = \text{softmax}(A^{(L)}).$$

2.5.3. Backpropagation Algorithm for Training Neural Networks on Images with Occlusion

The loss is calculated as:

$$(2.103) \quad \mathcal{E} = -\sum_{c=1}^C y_c \log \hat{y}_c.$$

During backpropagation the gradient values are computed with respect to the occluded image:

$$(2.104) \quad \frac{\partial \mathcal{E}}{\partial W^{(l)}} = \frac{\partial \mathcal{E}}{\partial A^{(l)}} * A^{(l-1)}$$

$$(2.105) \quad \frac{\partial \mathcal{E}}{\partial A^{(l)}} = W^{(l)} * \frac{\partial \mathcal{E}}{\partial A^{(l+1)}}.$$

Since

$$(2.106) \quad A^{(0)} = I \odot M,$$

the gradients are also masked:

$$(2.107) \quad \frac{\partial \mathcal{E}}{\partial I_M} = \frac{\partial \mathcal{E}}{\partial A^{(0)}} * M.$$

Augmenting images with partial occlusions used in CNN training increases network robustness to real cases of partially occluded objects.

2.6. Conclusions

1. CNNs are the primary multilayer neural networks used in computer vision, thanks to their ability to extract spatial features through weight sharing and local receptive fields.
2. The variety of layers (convolutional, pooling, fully connected, specialized) allows flexible network design adapted to specific tasks.

3. The backpropagation algorithm remains the fundamental training method for multilayer neural networks.
4. The choice of activation function, proper weight initialization, and regularization techniques such as dropout, batch normalization, and data augmentation are critical for the successful training of deep networks.
5. The appropriate choice of loss function depends on the nature of the task (classification, regression, metric learning) and has a direct impact on training quality.
6. Pre-trained models and fine-tuning via transfer learning are approaches that enable effective operation even with limited data.
7. In addition to CNNs, other architectures such as RNNs, autoencoders, and transformers expand the capabilities of deep learning to different domains and tasks.

CHAPTER 3. DETECTION AND RECOGNITION OF PARTIALLY VISIBLE MILITARY OBJECTS IN OPTICAL IMAGES USING NEURAL NETWORKS

A systematic analysis of the performance of pre-trained CNNs, implemented in MATLAB, is presented for the recognition of military equipment in optical images. Two interrelated experiments were conducted: evaluation of the baseline accuracy of the original networks for classifying military object images, and application of transfer learning to improve accuracy when classifying partially visible military objects.

3.1. Materials and Methods

3.1.1. Database and Data Preparation

A collection of 2748 images of military equipment distributed across eight classes was used: artillery, IFV (BMP), UAV (BPLA), armoured vehicle (bronemashina), APC (BTR), infantry, MLRS (RSZO), and tank. Some images were taken under real combat and training conditions, with the presence of multiple objects and cases of partial occlusion. The MATLAB ImageDatastore structure was used for data management, providing efficient loading and automatic labelling without the need to load all images into memory simultaneously.

3.1.2. Numerical Experimentation Environment — MATLAB Deep Learning Toolbox

In MATLAB, the Deep Learning Toolbox™ provides functions, applications, and blocks for designing, training, simulating, and deploying multilayer neural networks, including CNNs. The included tools allow visualization and analysis of network performance, tracking of performance metrics, application of optimization techniques, and deployment of models on resource-constrained systems.

3.1.3. Pre-trained Convolutional Neural Networks in MATLAB

The MATLAB Deep Learning Toolbox provides 19 pre-trained CNN architectures — from classical (AlexNet, VGG-16/19, GoogLeNet) to modern (ResNet, DenseNet, EfficientNetB0, NASNet) and compact (MobileNet-v2, ShuffleNet, SqueezeNet). Most are trained on the ImageNet dataset (1.2 M images, 1,000 classes); only DarkNet-53 was pre-trained on the COCO dataset. The number of CNN parameters ranges from 1 to 143 million.

3.2. First Experiment — Evaluation of Pre-trained Network Accuracy

3.2.1. Experimental Methodology

In the first experiment, the mean accuracy, minimum, maximum, and mean confidence of all 19 pre-trained networks available in MATLAB were analyzed on the 2,748-image database of military equipment, without additional training. Such systematic evaluation allows the most suitable models for the specific task to be selected and their use in real military applications to be optimized.

3.2.2. Classification Results

All 19 networks were tested on the full image database without additional training. Mean confidence varied between 54.56 % (EfficientNetB0) and 73.64 % (DenseNet201).

Table 3.3. Comparative analysis of results from the experiment with original pre-trained networks.

Model	Image Count	Mean Confidence [%]	Standard Deviation [%]	Min Confidence [%]	Max Confidence [%]	Detected Classes [%]
ALEXNET	2748	62.71	26.47	4.23	100	192
VGG16	2748	57.82	27.03	4.82	99.94	156
VGG19	2748	59.27	27.62	3.15	99.99	143
GoogLeNet	2748	68.47	25.37	8.11	100.00	138
InceptionV3	2748	61.82	25.75	3.67	99.54	120
ResNet18	2748	62.10	27.71	4.84	99.99	154
ResNet50	2748	66.70	27.56	4.42	100.00	156
ResNet101	2748	70.90	26.62	7.30	100.00	136
DenseNet201	2748	73.64	24.33	7.98	99.99	121
MobileNetV2	2748	55.45	27.70	3.05	99.93	162
NASNetLarge	2748	70.37	21.36	8.79	95.32	105
NASNetMobile	2748	63.03	25.09	4.46	99.17	126
ShuffleNet	2748	58.52	27.47	4.55	99.97	175
SqueezeNet	2748	62.20	26.22	6.61	99.99	198
Inception ResNetV2	2748	68.61	23.36	10.69	98.30	103
Xception	2748	56.56	27.42	3.10	99.98	116
DarkNet19	2748	58.18	29.43	3.19	99.97	159
DarkNet53	2748	70.25	26.39	6.57	100.00	128
EfficientNetB0	2748	54.56	26.01	4.62	98.54	122

3.2.3. Analysis of Most Frequently Recognized Classes

All recognized classes were military objects or machines that can be associated with military objects, or large agricultural machines that under poor-quality images can be confused with military objects (Fig. 3.1).

DarkNet53 is the only network trained on the COCO image collection. Its statistical parameters from the experiment are comparable to those of the other 18 networks. With a well-structured training approach, comparable results can be achieved with different image collections.

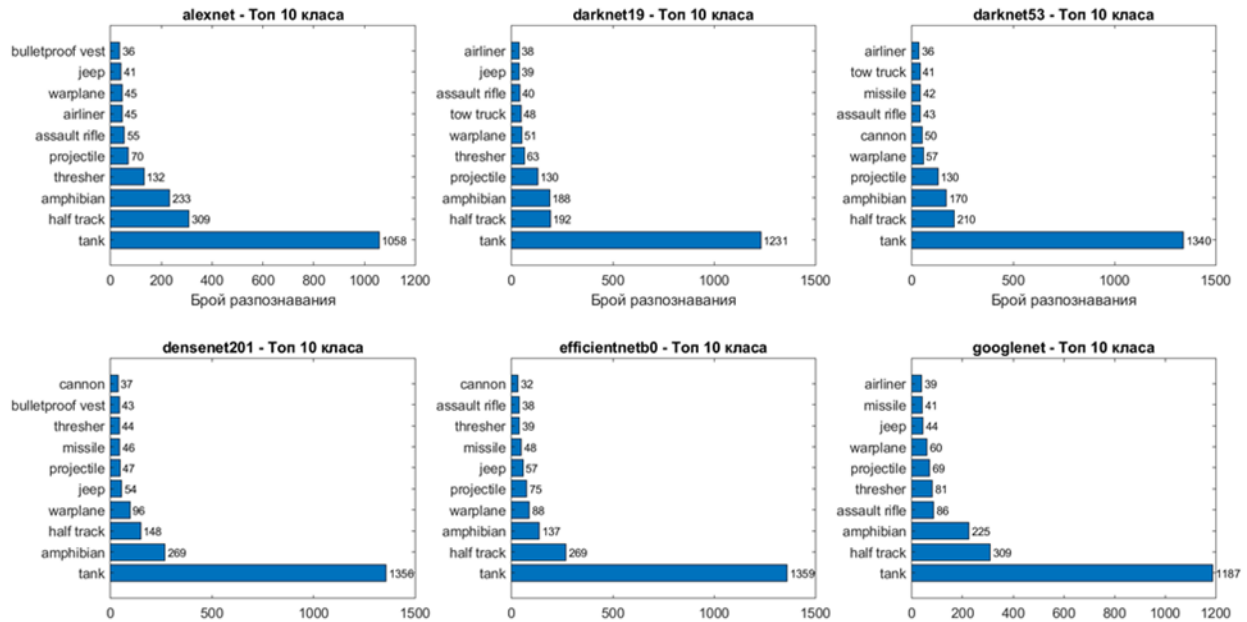


Fig. 3.1. Ten most frequently recognized classes for 6 selected models with different architectures.

3.2.4. Accuracy Analysis

In the data presented in Table 3.5, the accuracy of the networks is of the same order — no more than 14.41 %.

Table 3.5. Accuracy for six pre-trained networks with different architectures and data for the recognized class "tank".

Net	AlexNet	GoogLeNet	ResNet18	ResNet50	DarkNet19	EfficientNetB0	Mean
Accuracy	12.10 %	14.26 %	13.47 %	14.02 %	14.06 %	14.41 %	13.72 %
Detected Class: "Tank" (N, %)	1058 38.55 %	1187 43.19 %	1227 44.65 %	1312 47.74 %	1231 44.79 %	1359 49.45 %	1229 44.72 %

The images labelled as tanks in the collection number are 415 and they represent only 15.10 % of all images. The analysis shows that self-propelled artillery are often classified as "tank" due to physical and visual similarities, as well as APCs, IFVs, and armored vehicles with armament. Many of the "learned" characteristics

by the networks between military objects and certain heavy construction and agricultural machinery are also similar.

3.3. Second Experiment — Transfer Learning

3.3.1. Transfer Learning Experimental Methodology

Transfer learning was applied to six selected architectures (AlexNet, GoogLeNet, ResNet18, ResNet50, DarkNet19, EfficientNetB0) by replacing the output layers and additionally training only the new layers with the remaining weights "frozen".

3.3.2. Reclassification and Evaluation

Evaluation includes comparison of per class accuracy before and after additional training of the networks.

3.3.3. Approach for Simulating Partial Occlusion

To evaluate robustness under occlusions, the original database was augmented by systematically occluding images with black rectangles: 25 % occlusion (one quadrant) and 50 % occlusion (two quadrants), realized in 4 orientations. The approach increases the volume of the military object image database nine fold and enables controlled testing at different degrees of visibility.

3.3.4. Results of the Transfer Learning Experiment

Transfer learning led to significant accuracy improvements for all networks (Table 3.6 and Fig. 3.5). The highest accuracy on the original images was achieved by EfficientNetB0 — 95.60 %, followed by AlexNet (93.17 %) and ResNet50 (93.13 %). An exception was DarkNet19 with a weaker improvement of 40.10 %; negative transfer for the class "tank" was observed in GoogLeNet and DarkNet19 (−20.7 % and −19.5 %, respectively).

Table 3.6. Comparative analysis of overall accuracy results for classifying images from the original database, between original pre-trained and retrained (transfer learning) neural networks.

Net	Original, %	Retrained, %	Difference, %
AlexNet	12.10	93.17	81.07
GoogLeNet	14.26	84.80	70.54
ResNet18	13.47	90.85	77.38
ResNet50	14.02	93.13	79.10
DarkNet19	14.06	54.16	40.10
EfficientNet-b0	14.41	95.60	81.19



Fig. 3.5. Classification results of military objects from the retrained EfficientNetB0 network.

3.3.5. Results for Partially Visible Objects

When classifying augmented images (25 % and 50 % occlusion), the accuracy of the retrained networks dropped to the range 63 – 80 %, with EfficientNetB0 achieving 80.31 %. The original pre-trained networks performed at 8 – 11 % accuracy on the augmented images (Table 3.9).

Table 3.9. Comparative analysis of overall accuracy results for classifying partially occluded augmented images from retrained (transfer learning) neural networks.

Net	Original, %	Retrained, %	Difference, %
AlexNet	7.98	68.54	60.57
GoogLeNet	9.56	63.04	53.49
ResNet18	9.50	69.18	59.68
ResNet50	10.27	72.48	62.22
DarkNet19	10.18	32.19	22.02
EfficientNet-b0	10.93	80.31	69.38

3.3.6. Results of Transfer Learning with Augmented Images

Including the augmented images in the training process led to a substantial additional improvement of approximately 24 %. The EfficientNetB0 network (Net3), trained on the augmented database, achieved an overall accuracy of 99.25 %

for recognizing the class "tank" at 50 % occlusion, compared to 75.18 % for the same network without augmentation in training (Table 3.12). Results were stable across all occlusion variants, including with the lower or upper half-frame occluded.

Table 3.12. Comparative analysis of per-class accuracy results for classifying images from the augmented database for three differently trained EfficientNetB0 networks.

Class Name	Differently Trained EfficientNet-b0 Networks		
	Net1 (Orig)	Net2 (No Aug)	Net3 (Aug)
artilleriya	0.00 %	69.70 %	97.67 %
bmp	0.00 %	76.78 %	95.40 %
bpla	0.00 %	95.28 %	98.03 %
bronemashina	0.00 %	81.12 %	98.10 %
btr	0.00 %	75.05 %	96.45 %
pehota	0.00 %	91.96 %	96.76 %
rszo	0.00 %	81.94 %	98.71 %
tank	67.07 %	75.18 %	99.25 %

3.4. Conclusions

1. Despite significant differences in architectures and parameter counts (from 1 to 143 million), original pre-trained neural networks achieve similar baseline accuracy (12– 15 %) on military object recognition. The analysis shows that pre-trained networks tend to associate military objects with semantically related categories such as tank, half-track, amphibian, and warplane, given the visual similarity of military equipment.

1. Despite significant differences in architectures and parameter counts (from 1 to 143 million), original pre-trained neural networks achieve similar baseline accuracy (12 – 15 %) on military object recognition. The analysis shows that pre-trained networks tend to associate military objects with semantically related categories such as tank, half-track, amphibian, and warplane, given the visual similarity of military equipment.

2. Transfer learning is an effective method for adapting existing models to the specific task of military object recognition, improving accuracy by up to 81.19 % compared to original networks. Best results on original images were demonstrated by EfficientNetB0 with 95.60 % accuracy after transfer learning, followed by AlexNet (93.17 %) and ResNet50 (93.13 %) (Table 3.6).

3. For partially visible object recognition, the accuracy of pre-trained networks drops to 8 – 11 %, and for retrained networks without augmentation, it ranges from 63 % to 80 % (Table 3.9).

4. Training neural networks with augmented images containing partially occluded objects leads to substantial accuracy improvements — for example, EfficientNetB0 achieves 99.25 % accuracy for recognizing partially visible objects of class "tank" (Tables 3.12 and 3.13).

5. The analysis of experimental results shows the need for a specialized convolutional neural architecture designed for recognizing military objects in optical images in the presence of partial occlusion, poor visibility, or multiple objects in the frame.

CHAPTER 4. DESIGN OF A CONVOLUTIONAL NEURAL NETWORK FOR DETECTION AND RECOGNITION OF PARTIALLY VISIBLE MILITARY OBJECTS

Chapter 4 presents the design, systematic investigation, and optimization of a specialized convolutional neural architecture for recognizing partially visible military objects. The main goal is the creation of a compact model with minimal architectural complexity, applicable in systems with limited computational resources — UAV onboard computers, ground observation stations, and embedded platforms.

4.1. Design of Multilayer Convolutional Neural Networks for the Task

To ensure comparability, all designed architectures operate with the same input image format ($224 \times 224 \times 3$) and identical output structure of 8 classes. Configurations with 0 to 14 convolutional blocks were investigated, with systematic variation of depth enabling the study of the effect of architectural complexity on accuracy and robustness under occlusion.

4.1.1. Baseline Convolutional Neural Network with Four Convolutional Blocks

The baseline architecture consists of four sequential convolutional blocks, two fully connected blocks, and an output layer. Each convolutional block includes a convolutional layer (3×3 filters), a batch normalization layer, an activation layer (ReLU), and a pooling layer with a (2×2) mask. The number of filters increases progressively: $8 \rightarrow 16 \rightarrow 32 \rightarrow 64$, and the spatial resolution decreases: $224 \times 224 \times 3 \rightarrow 112 \times 112 \times 8 \rightarrow 56 \times 56 \times 16 \rightarrow 28 \times 28 \times 32 \rightarrow 14 \times 14 \times 64$. The fully connected blocks contain 1024 and 512 neurons respectively, each with ReLU activation and dropout ($p = 0.4$). The output layer has 8 neurons, followed by a Softmax activation layer and a classification layer. The total number of layers is 26.

The choice of 4 convolutional blocks is theoretically justified: the first layer extracts low-level features (contour lines, textures), the second — shape fragments and details of military equipment, the third — structural representations of components (turret, gun, hull), and the fourth integrates this information into abstract patterns capable of compensating for partial visibility through characteristic combinations of visible features.

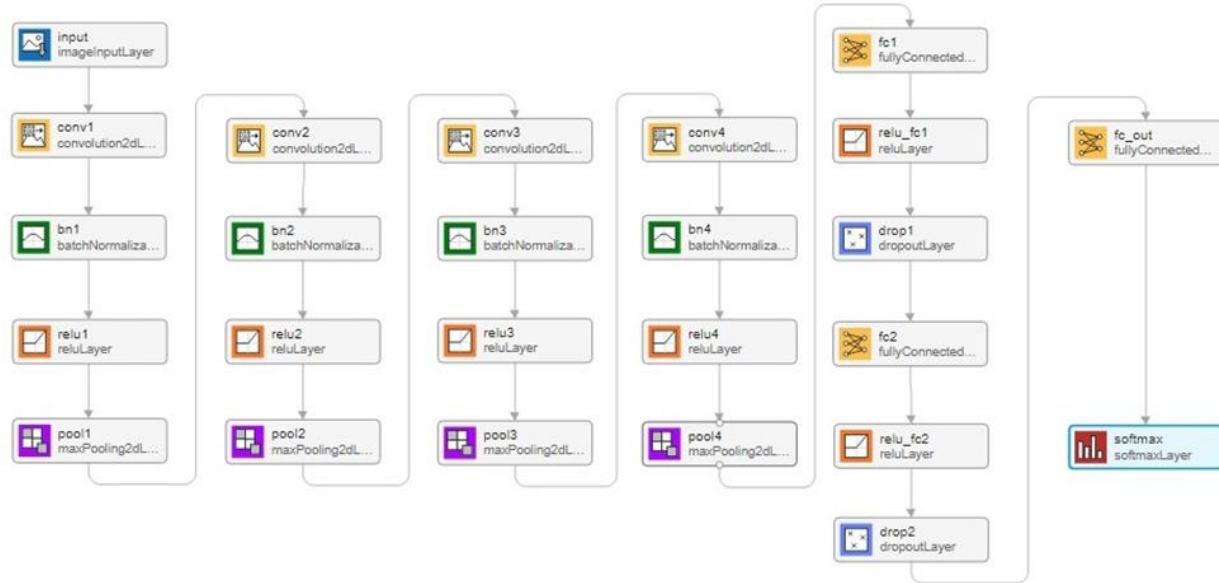


Fig. 4.2. Four-block CNN (without classification layer).

4.1.2. Modified Architectures

For systematic investigation of the effect of architectural complexity, the baseline network was extended in a series of modified variants with 3, 5, 6, 7, 8, 9, 10, 11, 12, and 14 convolutional blocks. Additionally, for specific configurations, variants with different convolutional filter sizes were tested while maintaining a fixed number of layers.

4.2. Training Methodology and Evaluation (Subsections 4.2.1–4.2.3)

Training and evaluation were structured in three sequential experimental stages, each using the same architectures with different volumes and quality of training data:

Stage 1 — training on the original collection of 2,748 images;

Stage 2 — training on sequentially augmented sets with geometric and photometric transformations ($\times 2$, $\times 4$, $\times 6$, $\times 8$ increase) without introducing partial occlusion in the training data;

Stage 3 — training on a specially constructed database with controlled partial occlusion up to 50 % (Chapter 3, Section 3.3), simulating real occlusion scenarios.

4.3. Experimental Results

4.3.1. General Testing Approach and Results Analysis

Evaluation of all architectures was performed on a unified test collection with partially occluded images, including originals and variants with 25% and 50% occlusion in four orientations. The registered metrics include mean classification accuracy, distribution of errors per class, accuracy degradation compared to original, and spatial asymmetry for occlusion.

4.3.2. Stage 1 Experimental Results

When training only on the original non-augmented collection, accuracy grew with the number of convolutional blocks up to a certain threshold. CNN-12 achieved the highest training accuracy (50.62 %), but with a gap of approximately 47 % between the training and validation curves — a clear sign of overfitting. Shallower architectures (≤ 7 blocks) demonstrated limited classification capability under partial visibility, while deeper ones showed unstable results. Mean accuracy on the occlusion-augmented collection ranged from 31 % (CNN-3) to 67 % (CNN-12).

4.3.3. Stage 2 Experimental Results

Increasing the training data volume led to systematic accuracy improvements for all architectures. At $\times 8$ augmentation, CNN-12 achieved 95.52 % training accuracy and 74.43 % mean accuracy on partially occluded images. Additional complexity did not yield a proportional effect — a "plateau" was observed at architectures with more than 7 – 8 blocks. The maximum average accuracy achieved with partial visibility in Stage 2 remained below 75 % for all architectures, confirming that standard augmentation alone is insufficient.

4.3.4. Stage 3 Experimental Results

Including controlled occlusion in the training data led to a qualitative change in results. Achieved accuracy values on partially visible images increased significantly for all architectures (Table 4.8). A clearly expressed maximum was observed at architectures with 6 – 8 convolutional blocks, after which further depth increase does not yield systematic improvement.

Table 4.8. Comparative analysis of training accuracy and mean classification accuracy on a partial-occlusion database depending on the number of convolutional blocks.

CNN — No. of convolutional blocks	3 blocks	4 blocks	5 blocks	6 blocks	7 blocks	8 blocks	9 blocks	10 blocks	11 blocks	12 blocks	14 blocks
Training accuracy on augmented collection, %	78.14	86.58	85.99	92.08	93.56	94.17	93.82	93.01	92.22	95.42	92.44
Mean accuracy on full partial-occlusion collection, %	82.41	90.92	91.87	96.97	97.87	97.82	97.71	97.05	96.87	98.21	96.88

4.4. Selection of the Optimal Architecture

4.4.1. Comparative Analysis for Optimal Architecture Selection

The selection of the optimal architecture is based on three criteria: (1) mean classification accuracy greater than 95 % for partially visible images; (2) achieving this accuracy with the smallest possible number of convolutional blocks; (3) stability — minimal difference between training accuracy and testing accuracy on occluded images.

Analysis shows that architectures with 3, 4, and 5 convolutional blocks do not reach the 95 % threshold. At 6 convolutional blocks a significant jump is observed — accuracy reaches 96.97 % with stability of 4.89 %. Architectures with 7 – 14 blocks achieve similar or slightly higher values, but additional improvement is not proportional to increased complexity.

The CNN-6 architecture with 6 convolutional blocks (34 neural layers) was identified as the optimal compromise between effectiveness and architectural complexity — the first configuration at which high and stable accuracy is achieved when classifying partially visible military objects (Fig. 4.9).



Fig. 4.9. Diagram of the implemented CNN with six convolutional blocks (without classification layer).

4.4.2. Detailed Description and Optimization of the Selected CNN-6 Architecture

After selecting the architecture, a two-stage internal optimization was conducted without changing the number of layers.

Optimization of the number of convolutional filters: With a fixed 3×3 size, different configurations of filter count and sequence were tested. Results show that the configuration $32 \rightarrow 64 \rightarrow 128 \rightarrow 256 \rightarrow 512 \rightarrow 512$ achieved the highest mean accuracy — 98.92 % under partial occlusion, representing an improvement of 1.95 % over the baseline configuration $8 \rightarrow 16 \rightarrow 32 \rightarrow 64 \rightarrow 128 \rightarrow 256$.

Optimization of the first convolutional filter size: With a fixed filter configuration, the effect of changing the first layer filter size (3×3 , 4×4 , 5×5 , 6×6 , 7×7) was tested. Results show that a 5×5 size gave a slight improvement (98.97 %), interpreted as a better balance between local detail extraction and wider spatial

context coverage. At larger sizes (6×6 , 7×7) accuracy slightly decreased due to loss of fine local features.

The final optimized CNN-6 architecture includes: input layer ($224 \times 224 \times 3$), six convolutional blocks with 5×5 filters in the first and 3×3 in the others, filter count sequence $32 \rightarrow 64 \rightarrow 128 \rightarrow 256 \rightarrow 512 \rightarrow 512$, two fully connected blocks (1024 and 512 neurons), Softmax and classification layer.

4.4.3. Results from the Detailed Analysis of the Optimized Six-Block CNN

Full statistical analysis by class and occlusion variant is provided (Table 4.12).

Table 4.12. Accuracy results for classifying occluded images by class and variant — optimized CNN-6.

Variants	Class and Variant Accuracy, %							
	artilleriya	bmp	bpla	bronemashina	btr	pehota	rszo	tank
Original	100	99,40	99.26	99,12	99,33	99,51	100	100
Quad1_TopL	100	99,60	99.26	99,12	98,65	99,03	100	100
Quad2_TopR	100	99,00	99.26	99,41	100	99,51	100	100
Quad3_BottomL	100	99,40	99.26	99,41	98,65	99,03	100	100
Quad4_BottomR	100	98,80	99.26	99,41	99,33	99,51	100	100
Half1_Top	98,84	98,21	98.15	99,41	97,64	99,03	99,71	99,76
Half2_Bottom	97,67	95,02	95.94	97,35	94,61	97,57	99,13	94,70
Half3_Left	99,42	99,20	98.15	98,53	97,64	98,54	100	99,28
Half4_Right	98,84	98,01	99.26	98,53	98,32	99,51	100	100
Mean accuracy	99,42	98,52	98.65	98,92	98,24	99,03	99,87	99,30
Minimum	97,67	95,02	95.94	97,35	94,61	97,57	99,13	94,70
Maximum	100	99,60	99.26	99,41	100	99,51	100	100
Standard deviation	0,82	1,42	1,12	0,69	1,57	0,64	0,09	1,74

Testing on an additional database with new occlusion variants (5 %, 10 %, 15 %, 20 %, 75 %) shows that the network successfully generalizes to variants not seen during training. At 5 % occlusion accuracy is approximately 99.5 %, at 10 % approximately 99 %, at 15 % about 94 – 97 %, at 20 % central occlusion 85.66 %, and at 75 % occlusion 83 – 93 % depending on orientation (Fig. 4.15).

These results confirm that accuracy is a nonlinear function of the visibility coefficient η and also has a qualitative dimension — it depends not only on the percentage of occlusion, but also on which specific part of the object is occluded.

Additional testing with impulse (salt & pepper) noise at densities 0.0 – 1.0 (Table 4.14) shows that the network maintains acceptable classification accuracy up to density 0.4 (83.51 %) (Fig. 4.16).

It should be noted that the occlusion operator introduced in the mathematical model in Chapter 2 here takes random values rather than only 0 and 1. In the forward propagation through the neural network this leads to random activations and

disruption of the correlation patterns stored during training, which affects the classification accuracy.

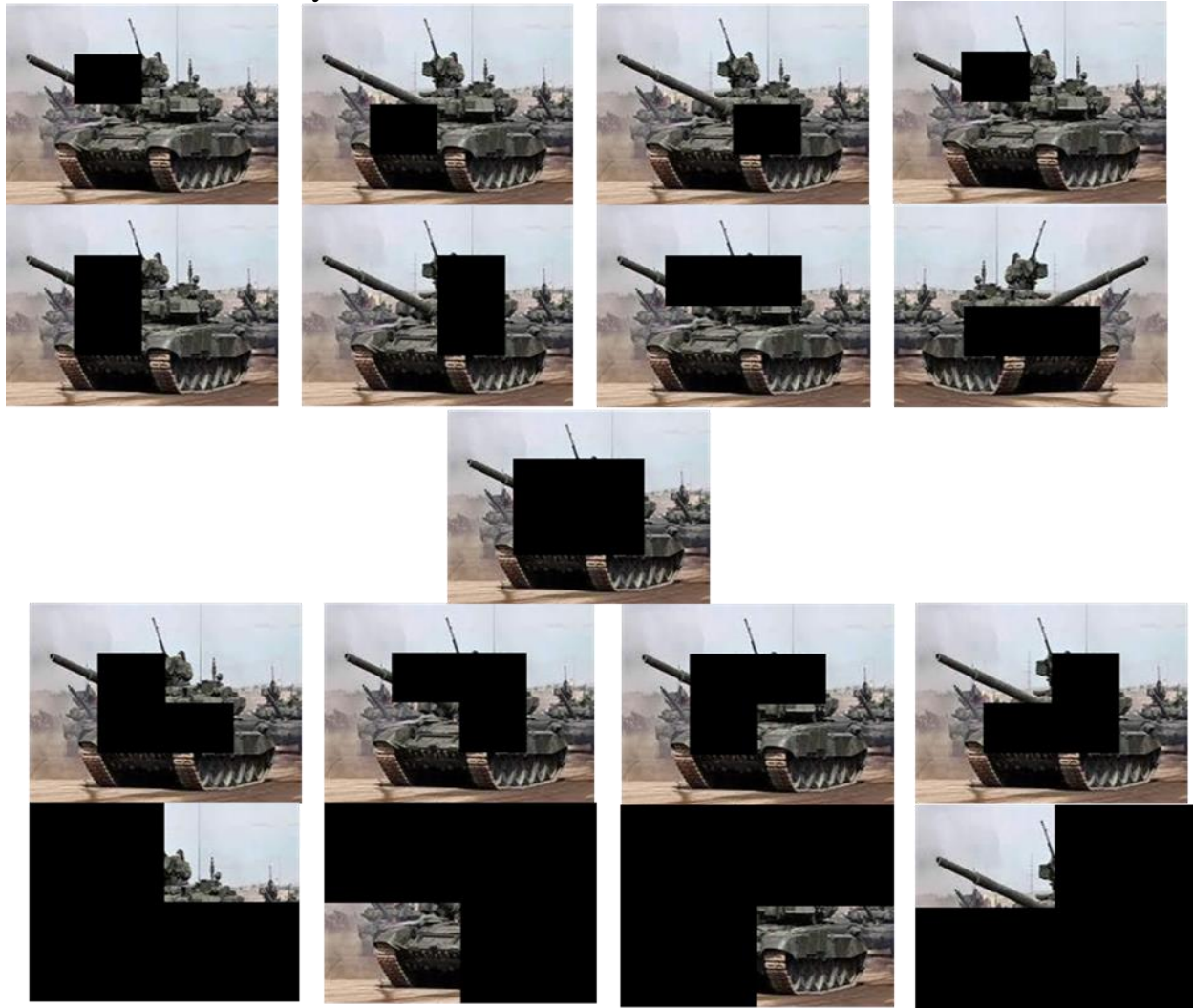


Fig. 4.15. Additional occlusion variants.

Table 4.14. Comparative analysis of accuracy results for the optimized six-block CNN classifying images partially occluded by impulse noise.

Noise density parameter value	Classification accuracy (%)	Hardest variant to classify, name and accuracy (%)
0,0	99,57	Quad1_TopLeft (99,56)
0,1	98,15	Half1_Top (94.58)
0,2	93,71	Half1_Top (76.79)
0,3	88,42	Half1_Top (63.16)
0,4	83,51	Half1_Top (53.65)
0,5	79,44	Half1_Top (47.60)
0,6	76,14	Half1_Top (43.01)
0,7	73,43	Half1_Top (40.22)
0,8	71,17	Half1_Top (37.86)
0,9	69,30	Half2_Bottom(34.96)

1,0	67,97	Half2_Bottom(32.29)
-----	-------	---------------------



Fig. 4.16. Occlusion variants Quad2_TopRight for a "tank" class image implemented via the salt&pepper (impulse noise) function with different noise densities.

Figure 4.24 presents the input image and all pseudo-colour maps of the mean per-channel values of the last layers in the six blocks, overlaid on the input image. This is interpreted as local image regions that are important for classification (averaged features highlighted by each block). The visualization shows how the network's "attention" concentrates on different details of the input image and extracts information about complex structures.

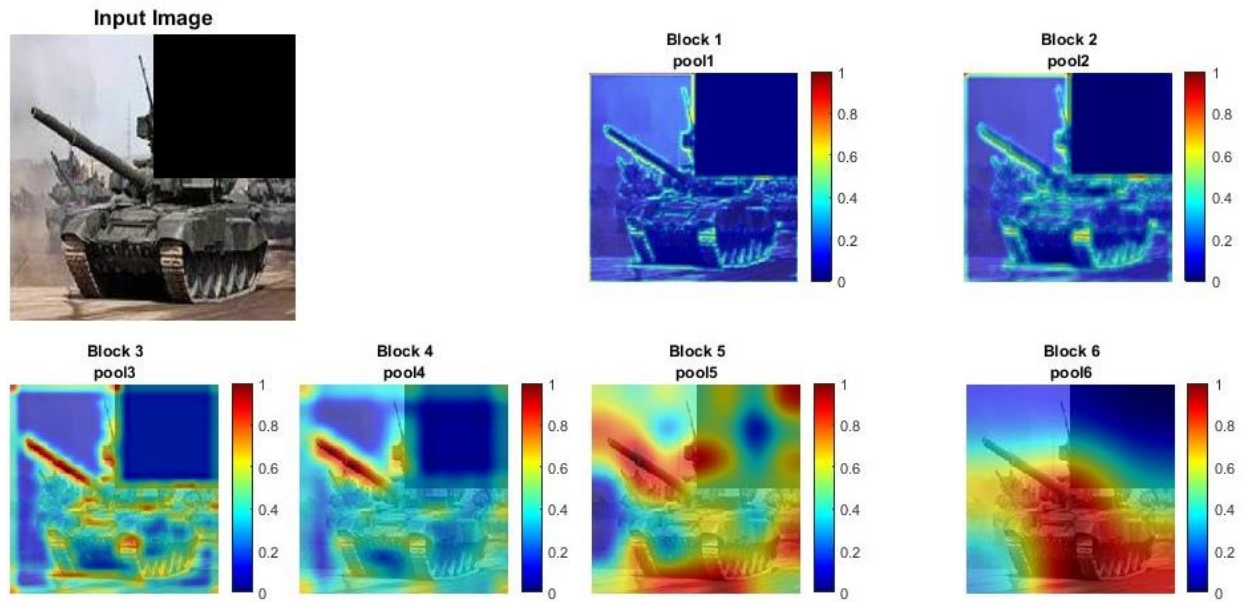


Fig. 4.24. All pseudo-color maps of the mean value per channel (feature maps) of the last (pooling) layers in the six blocks, overlaid on the input image.

4.5. Comparison Between the Retrained EfficientNet-B0 and CNN-6

4.5.1. Architectural Complexity

EfficientNet-B0 includes 237 layers and 5.3 million trainable parameters, while CNN-6 has 34 layers and 3.3 million parameters. CNN-6 was designed without the need for pre-training on large-scale datasets.

4.5.2. Comparison under Partial Occlusion

The comparison was performed on a low-performance processor (Intel Pentium B960), a mobile processor (Intel Core i3-5020U), and a server platform (2×Intel Xeon Gold 6530). Results (Table 4.15) show that the accuracy difference is less than 1%, while CNN-6 processing time is 2–3 times shorter on limited hardware. This confirms that a specialized, purpose-optimized architecture can achieve results comparable to established models at significantly lower computational cost.

Table 4.15. Comparative analysis of parameters of the retrained EfficientNet-B0 and CNN-6, obtained when running on three different hardware systems.

Parameters	EfficientNet-B0	CNN-6
Accuracy under occlusion	97.50%	96.97%
Model size	14,97 MB	12,84 MB
Time on Pentium (s)	0,358	0,109
Time on Core i3 (s)	0,195	0,107
Time on Xeon (s)	0,302	0,157

4.6. Detection of Partially Occluded Military Objects via Sliding Window Algorithm with the Trained Six-Block CNN

To extend CNN-6 to a localization task, a sliding window algorithm was implemented. The image is scanned in parts at different scales with a fixed step; each

region is resized and submitted for classification. Windows with confidence above a preset threshold form a two-dimensional detection map. Through filling and merging connected regions, a bounding box is extracted around the object. Additionally, a gradient-based activation map (Grad-CAM) was used for refining localization. This approach requires no additional training with annotated bounding boxes and is applicable with limited training data.

4.7. Cascade Model for Semantic Filtering and Recognition of Partially Visible Military Objects

For realistic scenarios with simultaneous presence of civilian and military objects, a cascade model was developed that combines the universal semantic power of EfficientNet-B0 (1,000 recognizable classes) and the precise domain-specific classification of CNN-6 for military equipment classes.

4.7.1. Theoretical Framework

The model implements hierarchical filtering in two main stages.

Stage 1A (semantic "gatekeeper"): EfficientNet-B0 classifies the input image. Upon recognizing a known civilian ImageNet class with confidence above a set threshold, the image is rejected as non-military.

Stage 1B (interval analysis): A 1,000-dimensional feature vector is extracted from EfficientNet-B0. The fundamental hypothesis is that this space contains implicit information about military objects, even though they are not explicitly defined as classes. Military classes are modelled through statistical intervals in the feature space. Only images with an interval estimate exceeding a predefined threshold τ are passed to Stage 2.

Stage 2: CNN-6 performs detailed classification among the eight military classes. When there is agreement between CNN-6 and the interval analysis, a final classification is accepted; otherwise a weighted decision is applied. For confirmed military objects, a localization procedure via sliding window is activated.

4.7.2. Mathematical Model for the Interval Classification Approach

Image I is transformed by a pre-trained neural network (feature extractor) into a feature space vector:

$$(4.1) \quad \mathbf{x} = f(I) \in R^d.$$

Military Object Class Model

For each class C_k , $k = 1, 2, \dots, K$, statistics are calculated:

$$(4.2) \quad X_{k=(x^{(1)}, x^{(2)}, \dots, x^{(N_k)})}.$$

Mean Vector

$$(4.3) \quad \mu_k = \frac{1}{N_k} \sum_{n=1}^{N_k} x^{(n)}.$$

Interval Model

For each coordinate $i \in \{1, 2, \dots, d\}$ an interval is defined:

$$(4.4) \quad [l_{k,i}, u_{k,i}],$$

$$(4.5) \quad l_{k,i} = \min_n x_i^{(n)},$$

$$(4.6) \quad u_{k,i} = \max_n x_i^{(n)}.$$

Interval Matching

For a new object x a function is defined:

$$(4.7) \quad M_k(x) = \sum_{i=1}^d w_{k,i} \delta_{k,i}(x),$$

$$(4.8) \quad \delta_{k,i}(x) = \begin{cases} 1, & l_{k,i} \leq x_i \leq u_{k,i} \\ 0, & x_i < l_{k,i} \cup x_i > u_{k,i} \end{cases},$$

$$(4.9) \quad 0 \leq M_k(x) \leq 1, \quad w_{k,i} \geq 0, \quad \sum_{i=1}^d w_{k,i} = 1.$$

Distance to Class Centre

$$(4.10) \quad D_k(x) = \sqrt{\sum_{i=1}^d w_{k,i} (x_i - \mu_{k,i})^2}.$$

Normalized Score

$$(4.11) \quad S_k(x) = \frac{1}{1+D_k(x)},$$

$$(4.12) \quad 0 < S_k < 1.$$

Combined Score

$$(4.13) \quad F_k(x) = \alpha M_k(x) + (1 - \alpha) S_k(x), \quad 0 \leq \alpha \leq 1.$$

Classification or Rejection Rule

$$(4.14) \quad \hat{k} = \underset{k}{\operatorname{arg\,max}} F_k(x), \quad F_{\max}(x) = \max_k F_k(x).$$

For a given threshold $\tau \in \{0,1\}$, the final decision is formed as:

$$(4.15) \quad F_{\max}(x) < \tau \Rightarrow x \neq \{1,2,\dots,K\},$$

$$(4.16) \quad F_{\max}(x) > \tau \Rightarrow x = \hat{k}.$$

4.7.3–4.7.7. Experimental Evaluation and Analysis of Results

The cascade model was tested on 3047 images — 299 civilian and 2748 military. With thresholds 0.80/0.75/0.90 the results show:

- Stage 1A rejected 150 images (4.9 %) as containing civilian objects (Fig. 4.26);
- Stage 1B additionally rejected 111 images (3.6 %) (Fig. 4.27);
- 2,786 images (91.3 %) reached CNN-6;
- CNN-6 and the interval analysis agreed on 58.2 % of images;
- Final classification: 2699 military (88.6 %), 261 civilian (8.6 %), 87 indeterminate (2.9 %) images;
- A bounding box was applied for 99.9 % of classified military objects.



Fig. 4.26. Images classified as containing civilian objects.

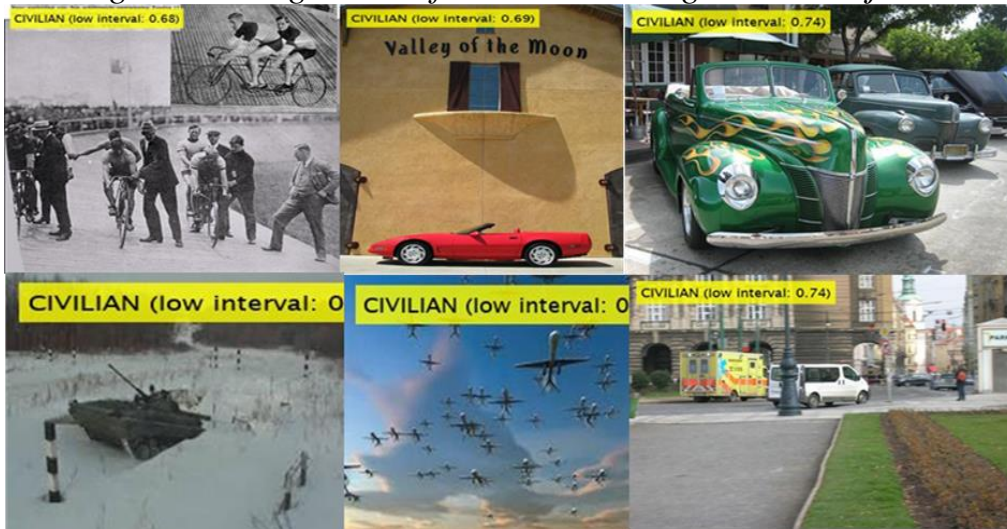


Fig. 4.27. Images with a maximum interval score below the preset threshold τ .

The classification accuracy of CNN-6 regarding military objects remains unchanged when integrated into the cascade, while the system-level image processing time remains within acceptable limits (1.87 s), even with the localization procedure enabled.

The contribution of the cascade approach lies in the analytical extension of a pre-trained network's functionality without changing learned weights and without additional training — through a deterministic mechanism using already available

semantic knowledge in the feature space. This turns large pre-trained architectures into universal semantic "sensors", extensible to new domains with minimal computational costs.

4.8. Conclusions

1. The staged experimental approach enables systematic investigation of the influence of architectural complexity (number of layers), training data volume, and presence of partial occlusion on the accuracy and robustness of CNNs for military object recognition.
2. Increasing the depth of the neural network architecture leads to improved classification accuracy only up to a certain threshold, beyond which further complexity brings no substantial benefits. This justifies the selection of a six-convolutional-block architecture (34 layers total) as the optimal choice in terms of accuracy and computational complexity.
3. Experiments with moderate training data augmentation show that increasing the training set volume improves model accuracy.
4. Training with controlled partial occlusion images leads to significant improvements in accuracy and robustness on realistic images (Tables 4.8, 4.12, and 4.14).
5. Comparative analysis with known retrained networks shows that the six-block CNN achieves comparable accuracy with significantly fewer trainable parameters, lower computational complexity, smaller model size, and shorter processing time (Table 4.15), making it suitable for implementation in mobile real-time systems.
6. The proposed CNN-6 architecture can also be used for military object detection tasks through implementation of a sliding window algorithm and analysis of local activations, without the need for additional training or annotated bounding boxes.
7. The developed cascade semantic filtering model combines a pre-trained universal neural network, analytical extension of the semantic space, and a specialized convolutional architecture, and represents a practically applicable approach for distinguishing between civilian and military objects.

3. CONCLUSION.

GENERAL CONCLUSIONS FROM THE DISSERTATION

1. As a result of the analytical review of existing object detection and recognition methods, it was established that modern solutions are oriented towards civilian applications and do not account for the specifics of military objects. The analysis revealed significant limitations of these methods in the presence of partial

occlusion, complex background, and multiple objects in the frame, which justified the need for developing and training a specialized neural network.

2. There is a need to develop a specialized database of optical images of military equipment including scenes from real combat and training conditions, partially occluded objects, and images with multiple objects. The database should provide the necessary representativeness for training, testing, and objective evaluation of the investigated neural architectures.

3. Based on the conducted analysis of existing neural architectures used in computer vision, a justified choice was established to utilize a multilayer CNN architecture in order to achieve the objective set in the dissertation.

4. The results from the systematic experimental evaluation of pre-trained well-known CNNs with the structured specialized database serve to assess their baseline classification capability and to identify their limitations regarding accuracy and robustness under partial occlusion.

5. The designed neural architecture with six convolutional blocks is characterized by improved accuracy when recognizing partially occluded military objects, as well as a better accuracy-to-computational-complexity ratio compared to the investigated pre-trained well-known neural networks.

6. The developed cascade semantic filtering model, in which a retrained neural network is used as a primary domain filter without full retraining on an extended class set, provides discrimination between civilian and military objects.

7. The comparative analysis between the optimized neural network and the selected pre-trained architectures demonstrates advantages of the proposed solutions in terms of accuracy, robustness under partial occlusion, and computational efficiency. The experimental results confirm that specialized architectural design and cascade processing organization are suitable for tasks related to military object recognition under real-world conditions.

In conclusion, the dissertation proposes a comprehensive and justified approach for detection and recognition of partially visible military objects in optical images. The presented architectures, methods, and algorithms are applicable in surveillance and decision support systems in the defence domain, as well as for future development of research in this direction.

4. CONTRIBUTIONS.

SCIENTIFIC AND APPLIED CONTRIBUTIONS OF THE DISSERTATION

1. An analytical review of existing artificial intelligence methods for object detection and recognition in optical images was conducted, including images of military equipment, as well as cases of partial occlusion of objects in the frame.

2. A specialized database of optical images of military equipment was structured, including scenes from real combat and training conditions, partially occluded objects, and images with multiple targets in the frame. The database is adapted for training, testing, and evaluation of neural network performance on military image recognition tasks (Chapter 3).

3. A systematic experimental evaluation of modern pre-trained well-known neural architectures was conducted on the specialized database, establishing their capabilities and limitations for recognizing military objects, including under partial occlusion conditions and increased scene complexity (Chapter 3).

4. An approach for training a specialized multilayer convolutional neural network architecture was proposed, designed for detection and recognition of partially visible military objects in optical images, with a focus on robustness under partial occlusion, complex background, and multiple objects — distinguishing it from standard solutions oriented towards civilian applications (Chapter 4).

5. A CNN was designed, implemented, and experimentally optimized in the MATLAB development environment, characterized by better robustness under partial occlusion and a better accuracy-to-computational-efficiency ratio compared to the investigated pre-trained architectures (Tables 4.9–4.14 and Figures 4.8–4.24).

6. A cascade semantic filtering model was proposed, in which a pre-trained neural network acts as a primary domain filter and the specialized neural network performs detailed classification of military objects. The model enables discrimination between military and civilian objects without the need for full retraining with an extended class set (Chapter 4, Section 4.8).

7. An algorithm for recognizing additional classes using features already extracted from a pre-trained neural network was developed. The algorithm uses statistical intervals and weighted evaluation in the feature space, without altering the original architecture and without changing the number of output classes (Chapter 4, Section 4.8.2, Appendix 2, Code 7).

8. A localization algorithm for military objects in images was integrated by implementing a sliding window, whereby the developed neural network is used as a specialized classifier for detecting and classifying partially occluded objects (Appendix 2, Codes 6 and 7).

9. A comparative analysis between the proposed optimized CNN and retrained well-known neural architectures was conducted, experimentally confirming the advantages of the proposed neural architecture and cascade approach in terms of accuracy, robustness under partial occlusion, and computational efficiency (Chapter 4, Sections 4.6 and 4.8).

APPLIED CONTRIBUTIONS OF THE DISSERTATION

1. The structured specialized database of optical images of military objects, including scenes with complex backgrounds, partial occlusion, different scales,

and multiple objects in the frame, was annotated and published for free access on the Kaggle platform for use by other researchers in testing neural networks.

2. A software prototype was developed in the MATLAB development environment, implementing the proposed neural architecture and algorithms, which can be deployed in surveillance and automated decision support systems in unmanned aerial vehicles for real-time operation (Appendix 2).
3. The developed programming algorithms were implemented as modular and extensible software components, allowing their adaptation to other image processing tasks, extension to additional object classes, use of other types of sensor data, or integration with existing software and hardware platforms.

LIST OF PUBLICATIONS

[1] Manov, L., & Kulev, H. (2023), Detection and recognition of partially visible objects in optical images using artificial intelligence techniques - mini review from 1970 to 2000, *Proceedings of International Scientific Conference "Defense Technologies" DefTech 2023* (pp. 444–450), Faculty of Artillery, Air Defense and Communication and Information Systems, National Military University, ISSN 2367-7902.

[2] Manov, L. E. (2024), Modern technologies with artificial intelligence for detecting military objects in optical images, *Proceedings of International Scientific Conference "Defense Technologies" DefTech 2024* (pp. 675–682), Faculty of Artillery, Air Defense and Communication and Information Systems, National Military University, ISSN 2815-4282.

[3] Manov, L. E. (2025). Modern artificial intelligence technologies for detection and recognition of partially visible military objects in optical images, *Proceedings of International Scientific Conference "Defense Technologies" DefTech 2025* (pp. 723–731), Faculty of Artillery, Air Defense and Communication and Information Systems, National Military University, ISSN 2815-4282.

[4] Manov, L. E. (2025), Modern artificial intelligence technologies for recognition military objects in optical images: a computer experiment with neural networks in MATLAB., *Proceedings of International Scientific Conference "Defense Technologies" DefTech 2025* (pp. 714–722). Faculty of Artillery, Air Defense and Communication and Information Systems, National Military University, ISSN 2815-4282.

[5] Manov, L. E. (2025), Modern artificial intelligence technologies for detecting and recognizing partially visible objects with military purpose in optical images: a computer experiment with transfer learning in MATLAB, *Proceedings of International Scientific Conference "Defense Technologies" DefTech 2025* (pp.732-741), Faculty of Artillery, Air Defense and Communication and Information Systems, National Military University, ISSN 2815-4282.