

УНИВЕРЗИТЕТ У НИШУ

ЕЛЕКТРОНСКИ ФАКУЛТЕТ

**ПРИМЕНА АУГМЕНТАЦИЈЕ СЛИКА У КЛАСИФИКАЦИЈИ РУКОМ ПИСАНИХ ЦИФАРА**

Семинарски рад

Студијски програм: Вештачка интелигенција и машинско учење

Модул:

Студент: Ментор:

Јована Стојановић, бр. инд. 1902 Проф. др Александар Станимировић

Ниш, август 2025. година

**САДРЖАЈ**

[1. **УВОД** 3](#_heading=h.7drk3txg9m9e)

[2. **ШТА ЈЕ АУГМЕНТАЦИЈА ПОДАТАКА?** 4](#_heading=h.9jp49mxr2809)

[2.1. Позадина 4](#_heading=h.k5d41za0kj6b)

[2.2. Текуће технике аугментације података и сродних концепата 4](#_heading=h.4ul5sebptxxl)

[2.3. Општи проблем аугментације података 5](#_heading=h.ozydh74h09bk)

[2.4. Примена и важност аугментације 6](#_heading=h.8vmyahac3qbc)

[3. **ТЕХНИКЕ АУГМЕНТАЦИЈЕ СЛИКА** 7](#_heading=h.9a214eshpzuy)

[3.1. Опште 7](#_heading=h.7obpb1xndzyp)

[3.2. Аугментације података засноване на основним манипулацијама слика 7](#_heading=h.8sefh83dy1kb)

[4. **АУГМЕНТАЦИЈА РУКОМ ПИСАНИХ ЦИФАРА** 16](#_heading=h.39xz1gg81ea7)

[4.1. Основе 16](#_heading=h.u6330ul5ukld)

[4.2. Преглед трансформација над цифрама 16](#_heading=h.3az51repc2r)

[4.3. Технологије, библиотеке и приступи 18](#_heading=h.sfgi0r42bygr)

[4.4. Ограничења и мане аугментације у препознавању рукописаних цифара 19](#_heading=h.8n5xa9qxhqsd)

[4.5. MNIST скуп података 20](#_heading=h.rg6p01s4b5k2)

[5. **ЗАКЉУЧАК** 22](#_heading=h.sngqfytd5rud)

[**ЛИТЕРАТУРА** 23](#_heading=h.ihv636)

ПРИМЕНА АУГМЕНТАЦИЈЕ СЛИКА У КЛАСИФИКАЦИЈИ РУКОМ ПИСАНИХ ЦИФАРА

**САЖЕТАК**

Аугментација података је широко коришћена техника у машинском и дубоком учењу, усмерена на превазилажење ограничења недовољних или неуравнотежених скупова података. Вештачким повећањем варијабилности података побољшава се генерализација модела и његова робустност.

Циљ овог рада је да представи концепт аугментације података, њену улогу као технике регуларизације и да прегледа најчешће методе које се примењују у пракси.

Рад се заснива на прегледу релевантне литературе, са фокусом на технике аугментације у рачунарском виду, као што су геометријске и фотометријске трансформације, убризгавање шума, насумично брисање, филтрирање и методе засноване на генеративним моделима. Поред тога, обрађене су и примене аугментације у другим доменима, укључујући обраду текста и звука.

Анализа показује да аугментација значајно смањује ризик од преобучавања, побољшава квалитет тренинг података и омогућава развој поузданијих модела чак и када су скупови података мали или неуравнотежени. Осим тога, напредни приступи као што је аугментација заснована на GAN моделима обезбеђују смислене синтетичке податке који додатно унапређују перформансе.

Аугментација података је кључни алат за изградњу ефикасних и прецизних система машинског учења. Њена ефективност зависи од правилног избора техника које су прилагођене домену и типу података, уз очување семантичког значења трансформација.

**Кључне речи:** аугментација података; обрада слика; убризгавање шума; насумично брисање; дубоко учење

APPLICATION OF IMAGE AUGMENTATION IN HANDWRITTEN DIGIT CLASSIFICATION

**ABSTRACT**

Data augmentation is a widely used technique in machine learning and deep learning, aimed at overcoming the limitations of insufficient or unbalanced datasets. By artificially increasing data variability, it improves model generalization and robustness.

The aim of this paper is to present the concept of data augmentation, its role as a regularization technique, and to review the most common methods applied in practice.

The study is based on a review of relevant literature, with a focus on augmentation techniques in computer vision, such as geometric and photometric transformations, noise injection, random erasing, filtering, and methods based on generative models. In addition, applications of augmentation in other domains, including text and audio processing, are addressed.

The analysis shows that augmentation significantly reduces the risk of overfitting, improves the quality of training data, and allows the development of more reliable models even when datasets are small or imbalanced. Furthermore, advanced approaches such as GAN-based augmentation provide meaningful synthetic data that further enhance performance.

Data augmentation is an essential tool for building efficient and accurate machine learning systems. Its effectiveness depends on the correct choice of techniques suitable for the domain and data type, ensuring that transformations preserve semantic meaning.

**Keywords:** data augmentation; image processing; noise injection; random erasing; deep learning

# УВОД

Развој алгоритама дубоког учења донео је значајан напредак у решавању сложених проблема, али квалитет и количина података и даље представљају кључни фактор успешности модела. Недовољно велики и разноврсни скупови података доводе до слабих генерализационих способности, што ограничава применљивост и поузданост модела. Са друге стране, прибављање великих, репрезентативних и означених скупова података често је скупо, споро или у неким доменима практично немогуће.

Овим радом обрађује се проблематика аугментације података, односно техника које омогућавају проширење и обогаћивање скупа за тренирање креирањем нових примера из постојећих. Испитивани материјал обухвата слике као најчешћи формат података у рачунарском виду, са посебним освртом на различите врсте геометријских и фотометријских трансформација, као и напредније методе засноване на генеративним моделима.

Област истраживања односи се на примену техника аугментације слика у контексту машинског учења и дубоких неуронских мрежа. Циљ рада је да се систематизују најзначајније технике аугментације, процени њихова употребна вредност у убради фотографија руком писаних цифара.

У раду су представљени следећи методи: геометријске трансформације, трансформације у колорном простору, филтрирање слика кернелима, мешање слика, случајно брисање, додавање шума, као и аугментација заснована на напредним моделима (GAN, *transfer learning*, *style transfer*).

Закључно, овај рад доприноси бољем разумевању значаја и улоге аугментације података у савременим системима за машинско учење, пружајући смернице за њихову примену и дајући основу за даља истраживања усмерена ка развоју ефикаснијих и прилагодљивијих алгоритама. Овај приступ не само да побољшава перформансе постојећих модела, већ и отвара пут ка иновативним решењима која могу бити применљива у бројним научним и индустријским областима.

# ШТА ЈЕ АУГМЕНТАЦИЈА ПОДАТАКА?

## 2.1. Позадина

Алгоритми машинског учења омогућавају рачунарским системима да уче како би решавали различите проблеме на основу посматраних података [1]. Упркос огромном напредку у алгоритмима дубогок учења у последњим годинама, добијање адекватних и репрезентативних података за тренирање остаје најважнији фактор који одређује перформансе модела машинског учења. Без обзира на иновације у дизајну модела и техникама тренирања, тренирање са малим и лоше репрезентованим скуповима података на крају ће довести до лоших способности генерализације. Са друге стране, богати, велики и репрезентативнији подаци готово увек гарантују добре перформансе, чак и при коришћењу мање софистицираних алгоритама.

За многе практичне задатке, прибављање довољно великог и репрезентативног скупа података остаје изазов. Студије показују да чак и данашњи велики скупови података као што су *ImageNet* и CIFAR-100 не поседују довољан ниво варијабилности да би гарантовали добре перформансе. Повећање робустности и тачности модела машинског учења и омогућавање њиховог доброг рада на малим, лоше репрезентованим подацима главни је циљ аугментације података. Ови се недостаци могу ублажити коришћењем одговарајућих техника аугментације података, при чему се додатни подаци за тренирање генеришу из постојећих података или се креирају од нуле коришћењем различитих техника.

Ефикасне стратегије аугментације података могу такође смањити потребу за сложеним алгоритмима и омогућити једноставнијим алгоритамским архитектурама модела да постигну високе способости генерализацје.

У случају рачунарског вида, подаци за трениранје обично су у облику слика или видео секвенци. У већини задатака, видео секвенца се третира као збирка слика по кадровима. Тродимензионални задаци у рачунарском виду обично се ослањају на податке у облику воксела[[1]](#footnote-0) и облака тачака.

Што су подаци за тренирање већи, разноврснији и репрезентативнији, то ће модел дубоког учења ефикасније радити на невиђеним подацима. Са једне стране, разноликост узорака за тренирање треба да буде довољна да модел може да се носи са различитим одступањима у изгледу слика, па чак и са бучним циљним инстанцама. Са друге стране, неопходно се да се квалитет не погорша толико да омета перформансе на нормалним сликама.

## 2.2. Текуће технике аугментације података и сродних концепата

Аугментација података је део ширег скупа техника регуларизације које имају за циљ побољшање перформанси модела. Методе регуларизације функционишу тако што у модел машинског учења уводе додатне информације, како би боље ухватиле општије особине проблема који се моделује. Најједноставнији и један од најраспрострањенијих приступа је увођење казненог члана у функцију губитка, којим се ограничавају вредности параметара модела. Ово омогућава да обучени модел има стабилнији и „глађи” одговор на различите варијације улазних података (у време тестирања).

Беч нормализација (*Batch normalization*) нормализује излазе различитих слојева како би се елиминисали ефекти унутрашњег коваријантног померања улазних података у дубљим слојевима. Регуларизација се такође може постићи мењањем ефективне величине и структуре модела дубоког учења (нпр. редуковањем модела - *model pruning*), променом броја параметара који се уче, или обрасцем синетичких конекција (нпр. *dropconnect*).

*Dropout* регуларизација, блиска методу *dropconnect*, подразумева уклањање одређених неурона током тренинга. Циљ ове методе је смањење сложености модела како би се спречило учење ирелевантних шаблона из података модела како би се спречило учење непотребних шаблона из података.

Друга група техника обухвата промену самог процеса тренирања, као што је рани прекид (*early stopping*), који зауставља тренирање када се перформансе на валидационом скупу почну погоршавати. Такође, смањивање тежина (*weight decay*) се користи као једноставан начин за контролу пренаглашеног учења (*overfitting*) у великим моделима.

Када је приступ великом скупу података ограничен, често се користе трансфер учење (*transfer learning*) и фино подешавање (*fine-tuning*). У овом случају, унапред обучени модел се прилагођава новом, али сличном задатку, чиме се избегава потреба за потпуним тренирањем од почетка. Овај приступ омогућава добре резултате и на малим скуповима података.

Још један чест приступ побољшању перформанси модела је развој бољих репрезентација обележја (*feature representations*) у подацима. У овом домену, посебно је важна област учења репрезентација (*representation learning*), која има за циљ да моделу омогући боље разумевање структуре података.

Аугментација података, којој је посвећен овај рад, представља скуп техника које побољшавају перформансе модела кроз измену или проширење тренинг скупа података. Ове технике се директно примењују на податке и омогућавају моделима да буду робуснији и способнији за генерализацију.

## 2.3. Општи проблем аугментације података

За дати скуп података , који се састоји од тренинг узорака са одговарајућим ознакама , задатак аугментације података је да се дефинишу трансформационе операције , које се могу применити на оригиналне узорке како би се створили додатни подаци за тренинг , без промене одговарајућих ознака .

Овакве врсте трансформација познате су као операције које чувају ознаке (*label-preserving transformations*) у литератури из области обраде слике и рачунарског вида. То значи да трансформације () резултују измењеним узорком који се и даље семантички може описати оригиналном ознаком .

Важно је напоменути да не постоји објективна мера „означености” (*labelness*), односно критеријум који прецизно дефинише шта значи да је ознака сачувана. Стандардна пракса је да се ослонимо на људску процену - сматра се да је ознака сачувана ако је трансформисана слика и даље препознатљива и задржава првобитно значење у људској перцепцији.

Генерална идеја је следећа: за тренинг узорке који припадају 𝑛-димензионалном просторy , трансформисани узорци такође треба да припадају истом простору:

() (, )

тј. трансформисаним узорцима се додељују исте ознаке као оригиналима.

Више различитих трансформационих операција може се применити ради повећања диверзитета података. Међутим, због ограничења у рачунарским ресурсима, трошкова и могућих негативних ефеката увођења нерелевантних (или лажних) података, препоручује се пажљив избор само најкориснијих облика аугментације.

## 2.4. Примена и важност аугментације

Најчешће се примењује у обради слика кроз две главне категорије трансформација: геометријске и фотометријске [2]. Поред њих, у новије време се користе и напредне технике као што су уметање шума, насумично одсецање делова слике (*random cropping*) или чак коришћење генеративних модела за синтетичко креирање потпуно нових примера.

Примена није ограничена само на рачунарски вид, већ и на друге домене машинског учења. У обради текста, аугментација се спроводи путем синонимских замена, парафразирања или превођења реченица на други језик и враћања назад (*back translation*). У анализи аудио сигнала, често се користи променa висине тона, додавање позадинске буке или промена брзине снимака. Сви ови приступи имају за циљ да створе варијабилност у подацима, а да се при томе не наруши њихово основно значење.

У здравству [3], омогућава обраду малих и осетљивих скупова медицинских слика - попут ретких болести - уз очување приватности. У финансијама, помаже у стварању синтетичких примера преваре ради боље детекције. У производњи, подржава откривање визуелних дефеката када стварних података нема довољно. У малопродаји, омогућава генерисање варијација слика производа за бољу класификацију по категоријама.

# ТЕХНИКЕ АУГМЕНТАЦИЈЕ СЛИКА

## 3.1. Опште

Најраније демонстрације које су показале ефикасност аугментације података потичу од једноставних трансформација као што су хоризонтално окретање, аугментације у колорном простору и случајно исецање. Ове трансформације обухватају многе инваријантности.

Аугментације које су обухваћене овим прегледом укључују: геометријске трансформације, трансформације у колорном простору, филтере језгра, мешање слика, случајно брисање, аугментацију у простору обележја (*feature space augmentation*), адверзаријално тренирање (*adversarial training*), аугментацију засновану на GAN[[2]](#footnote-1) мрежама и пренос неуронског стила (*neural style transfer*).

## 3.2. Аугментације података засноване на основним манипулацијама слика

### 

У овом одељку описане су различите аугментације засноване на геометријским трансформацијама и другим функцијама за обраду слика [4]. Ова класа аугментација се може карактерисати као лака за имплементацију. Разумевање ових трансформација представља корисну основу за даље истраживање техника аугментације података.

Такође ћемо описати различите геометријске трансформације у контексту њихове „безбедности“ примене. Под безбедношћу аугментације подразумева се вероватноћа да ће се ознака (*label*) сачувати након трансформације. На пример, ротације и хоризонтално окретање су углавном безбедне за задатке као што су препознавање мачака и паса у оквиру *ImageNet*-а, али нису безбедне за задатке препознавања цифара као што су разликовање 6 и 9.

Трансформације које не чувају ознаку могу потенцијално ојачати способност модела да изрази несигурност у предвиђању, што у одређеним сценаријима може бити пожељно. Међутим, да би се то постигло, потребно је користити рафинисане (вишезначне) ознаке након аугментације. На пример, ако се након трансформације ознака изрази као вектор [0.5, 0.5], модел може учити да боље процењује ниво сопствене сигурности. Ипак, конструисање рафинисаних ознака за сваку небезбедну трансформацију је веома захтеван и скуп процес са становишта обраде података.

Због тога је при планирању аугментације веома важно размотрити степен безбедности трансформације. Овај аспект је често зависан од домена примене, што представља изазов у развоју универзалних стратегија аугментације.

Не постоји ниједна функција за обраду слике која, на довољно високом нивоу изобличења, не може довести до промене ознаке. Ово јасно показује колико су аугментације специфичне за податке и колико је изазовно развити опште применљиве политике аугментације. Управо из тог разлога, овај аспект треба имати на уму приликом примене геометријских трансформација које су описане у наставку.

### 3.2.1. Окретање (*Flipping*)

Хоризонтално окретање (по хоризонталној оси) је знатно чешће у пракси од вертикалног окретања. Ова аугментација је једна од најједноставнијих за имплементацију и показала се као веома корисна на скуповима података као што су CIFAR-10 и *ImageNet*.

Међутим, у скуповима који укључују препознавање текста, као што су MNIST или SVHN, оваква трансформација не чува ознаке, односно није *label-preserving*, јер окретање може потпуно изменити значење цифре или карактера.

### 3.2.2. Простор боја (*Color space*)

Дигиталне слике се најчешће кодирају као тензори димензија (висина × ширина × број канала боја). Извођење аугментација у простору канала боја представља једну од стратегија која је веома практична за имплементацију.

Најједноставније аугментације у боји укључују издвајање једног канала, као што је *R* (црвени), *G* (зелени) или *B* (плави). Слика се лако може претворити у представу у једном каналу тако што се задржи само тај канал, док се преостала два замењују матрицама са нулама.

Поред тога, вредности у RGB каналу могу се једноставно мењати помоћу матричних операција, што омогућава лако повећање или смањење осветљености слике.

Напредније аугментације у простору боја заснивају се на извођењу хистограма боја који описује слику. Променом интензитета у хистограму могу се добити ефекти промене осветљења, слично као у класичним апликацијама за уређивање фотографија.

### 3.2.3. Исецање (*Cropping*)

Исецање слика представља практичан корак у обради слика са различитим висинама и ширинама, где се од сваке слике узима централни део (печ). Поред тога, случајно исецање (*random cropping*) може да пружи ефекат сличан транслацијама (померањима слике).

Разлика између случајног исецања и транслација је у томе што исецање смањује димензије улаза, на пример са (256 × 256) на (224 × 224), док транслације задржавају просторне димензије слике непромењеним.

У зависности од избора прага смањења димензија приликом исецања, ова трансформација можда неће сачувати ознаку (*label-preserving transformation*), те треба пажљиво приступити њеној примени.

### 3.2.4. Ротација (*Rotation*)

Ротационе аугментације се изводе окретањем слике улево или удесно око осе у распону од 1° до 359°. Безбедност примене ротације у великој мери зависи од величине угла ротације.

Мале ротације, на пример између 1° и 20° или −1° и −20°, могу бити корисне у задацима препознавања цифара као што је MNIST. Међутим, како угао ротације расте, ознака података се после трансформације све више мења, те ротација престаје да буде *label-preserving*.

### 3.2.5. Транслација (*Translation*)

Померање слика улево, удесно, горе или доле представља веома корисну трансформацију која помаже у избегавању позиционе пристрасности у подацима. На пример, ако су све слике у скупу података центриране, што је често случај у скупу за препознавање лица, модел би у том случају био тестиран само на савршено центрираним сликама.

Када се оригинална слика помери у неком правцу, преостали простор може бити попуњен константним вредностима као што су 0 или 255, или пак случајним или Гаусовим шумом. Овај поступак пуњења (*padding*) омогућава да просторне димензије слике остану непромењене након аугментације.

### 3.2.6. Увођење шума (*Noise Injection*)

Увођење шума подразумева додавање матрице случајних вредности, обично узетих из Гаусове расподеле. Ову технику су тестирали *Moreno-Barea[[3]](#footnote-2)* и сарадници на девет скупова података из UCI репозиторијума. Додавање шума на слике може помоћи конволуционим неуронским мрежама (CNN) да науче робусније карактеристике.

Геометријске трансформације су одлично решење за позиционе *bias*-е које се могу јавити у скупу података за тренирање. Постоји много потенцијалних извора пристрасности који могу разликовати расподелу података за тренирање и тестирање. На пример, ако постоје позиционе пристрасности као што је у скупу за препознавање лица где је свака фотографија савршено центрирана, геометријске трансформације представљају изврсно решење.

Поред јаке способности да се отклоне позиционе пристрасности, геометријске трансформације су корисне и због своје једноставности имплементације. Постоји много библиотека за обраду слика које олакшавају операције као што су хоризонтално окретање и ротација.

Међутим, неке мане геометријских трансформација укључују повећану потрошњу меморије, трошкове израчунавања трансформација и додатно време за тренирање. Такође, неке трансформације попут транслације или случајног исецања морају се пажљиво посматрати како се не би измениле ознаке слика.

На крају, у многим доменима примене, као што је анализа медицинских слика, пристрасности које разликују тренинг и тест скупове података су сложеније од позиционих и транслационих варијанти. Због тога је обим примене геометријских трансформација релативно ограничен.

### 3.2.7. Трансформације у колорном простору (*Color Space Transformations*)

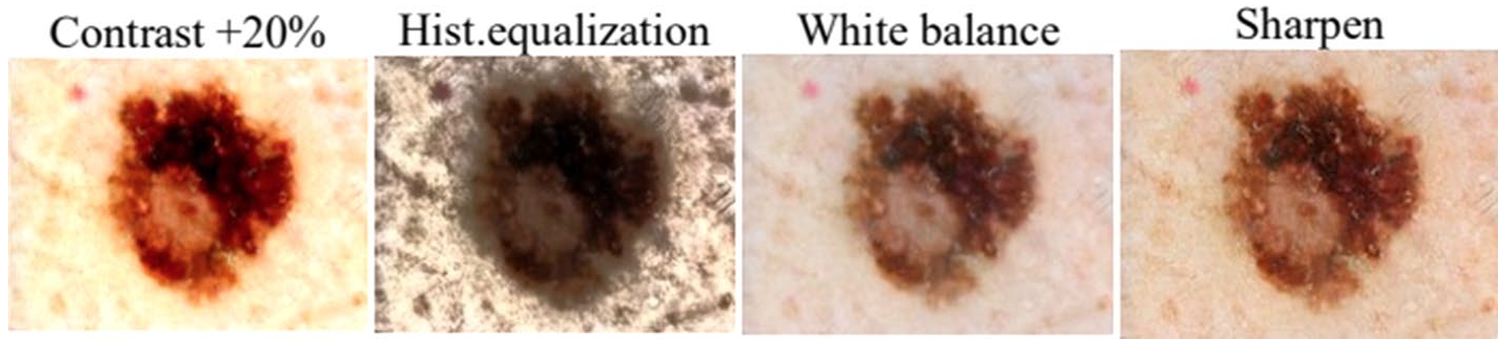
Слике се у дигиталном облику кодирају као три наслагане матрице димензија висина × ширина, од којих свака представља вредности пиксела за појединачну RGB боју (црвена, зелена, плава).

Осветљење је један од најчешћих изазова у задацима препознавања слика, због чега су трансформације у колорном простору - познате и као фотометријске трансформације - интуитивно ефикасне.

Једноставно решење за слике које су превише светле или тамне је да се пикселима дода или одузме константна вредност. Још једна лака техника је издвајање појединачних матрица из RGB канала. Друга трансформација подразумева ограничавање вредности пиксела на одређени минимум или максимум.

Многе трансформације у колорном простору потичу из апликација за уређивање слика. Вредности пиксела у сваком RGB каналу могу се агрегирати у хистограм боја, који се затим може модификовати ради примене различитих филтера који мењају визуелне карактеристике слике.

Трансформације у колорном простору нуде велику креативну слободу. Промена распореда боја у сликама може бити изузетно ефикасна у решавању проблема са неуједначеним осветљењем у тест скупу података (слике 1 и 2).



*Слика 1. Примери колорних аугментација у класификацији меланома*



*Слика 2. Примери колорних аугментација на уличној фотографији*

Скупови података слика могу се поједноставити тако што се RGB матрице конвертују у једну *grayscale* (црно-белу) слику. Овим се добијају слике мањих димензија (висина × ширина × 1), што резултира бржим рачунањем. Међутим, показано је да ова конверзија смањује тачност класификације. Забележен је пад тачности класификације од око 3% када се користе grayscale уместо RGB слика.

Осим разлике између RGB и *grayscale*, постоји више начина за представљање дигиталне боје, као што су HSV (*Hue*, *Saturation*, *Value*), YUV, CMY и други.

Слично као код геометријских трансформација, мане трансформација у колорном простору укључују већу потрошњу меморије, рачунске ресурсе и време тренирања. Поред тога, трансформације боја могу уклонити важне информације о боји, па самим тим нису увек *label-preserving* трансформације.

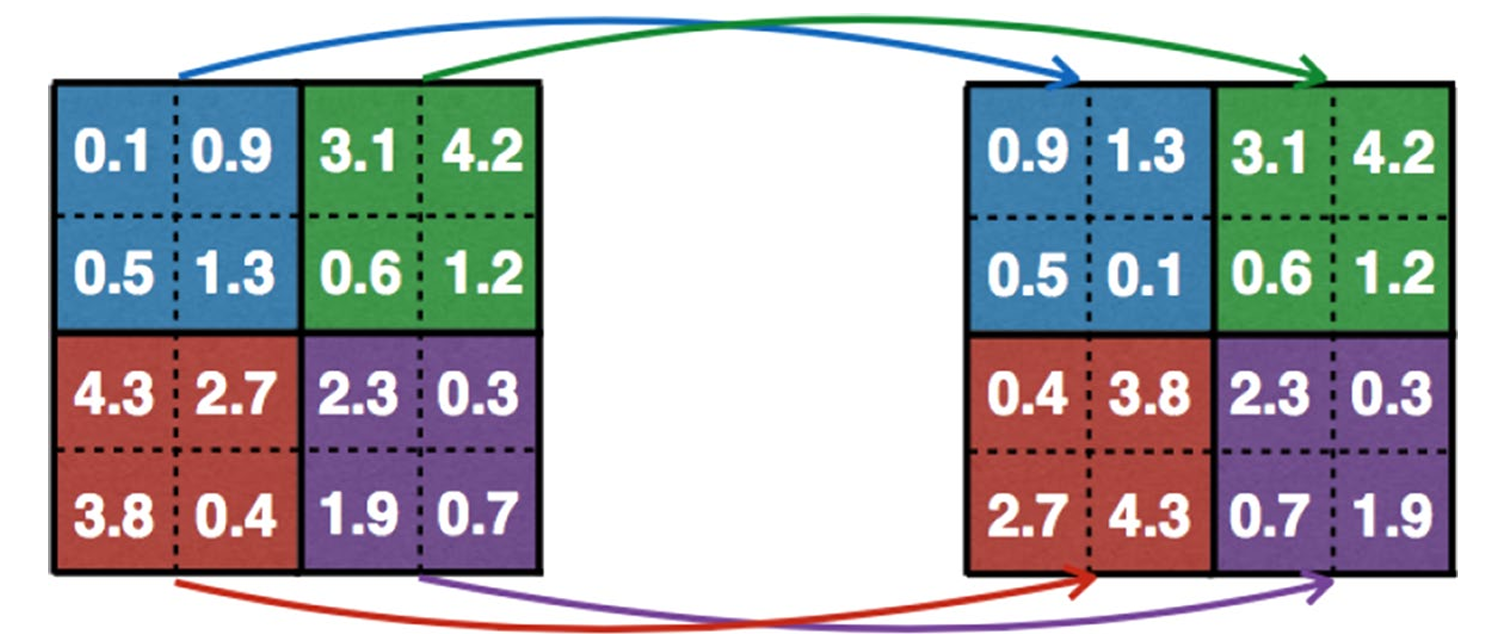
На пример, ако се вредности пиксела смање да би се симулирало тамније окружење, може постати немогуће уочити објекте на слици. Још један индиректан пример где се губи информација о ознаци је у анализи сентимента слика (*Image Sentiment Analysis*). У овом задатку, CNN модели покушавају да визуелно предвиде сентимент слике: врло негативан, негативан, неутралан, позитиван или врло позитиван. Један од кључних визуелних показатеља негативног сентимента је присуство крви, а тамноцрвена боја је одлучујућа за разликовање крви од воде или боје.

Ако трансформације боја толико измене простор боја да модел више не може разликовати крв од зелене боје, перформансе модела ће бити значајно ослабљене. Трансформације у простору боја могу елиминисати пристрасности у боји у корист просторних карактеристика, али за одређене задатке - као што је сентимент анализа или медицинска дијагностика - боја је кључна јединствена особина.

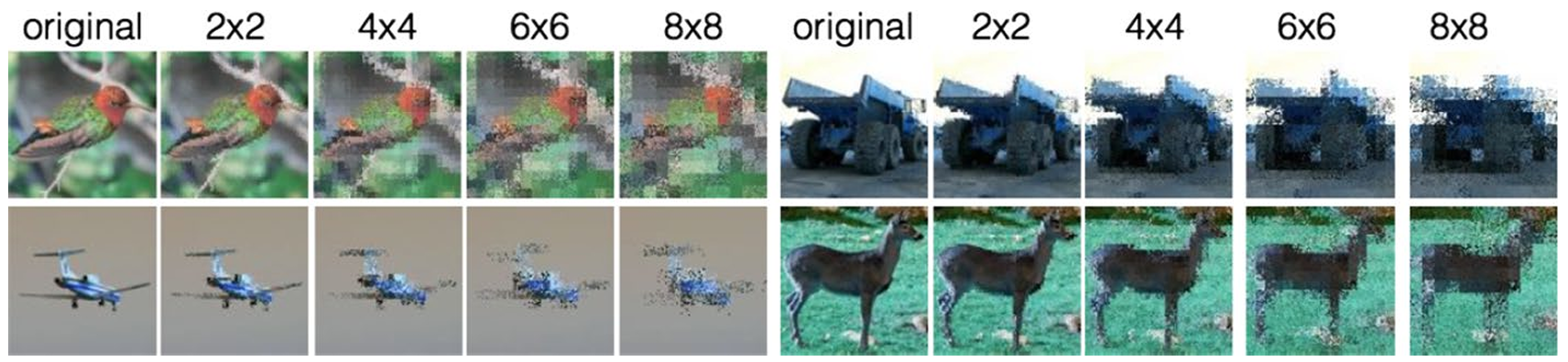
### 3.2.8. Кернел филтри

Кернел филтри су веома популарна техника у обради слика за изоштравање и замућивање слика. Ови филтери функционишу тако што се n × n матрица „провлачи“ преко слике, при чему се користи или Гаусово замућење, што резултира замућенијом сликом, или филтер за ивице високог контраста у вертикалном или хоризонталном правцу, што резултира оштријом сликом дуж ивица. Интуитивно, замућивање слика као техника аугментације података може довести до веће отпорности на замућење услед покрета током тестирања. Са друге стране, изоштравање слика у контексту аугментације података може резултирати бољим „хватањем“ детаља објеката од интереса.

Изоштравање и замућивање представљају класичне начине примене кернел филтера на слике. Истраживачи експериментишу са јединственим кернел филтером који насумично мења вредности пиксела у n × n прозору који се креће преко слике. Ову технику аугментације називају *PatchShuffle Regularization*. Експериментишући са различитим величинама филтера и вероватноћама за насумично мешање пиксела у сваком кораку, показали су ефикасност ове технике постигавши стопу грешке од 5,66% на CIFAR-10 датасету, у поређењу са 6,33% без примене *PatchShuffle Regularization*-а. Хиперпараметри који су довели до овог резултата били су филтри величине 2 × 2 и вероватноћа мешања од 0.05. Ови експерименти су изведени користећи *ResNet* архитектуру конволуционе неуронске мреже (слике 3 и 4).



*Слика 3. Примери примене технике PatchShuffle регуларизације*

**

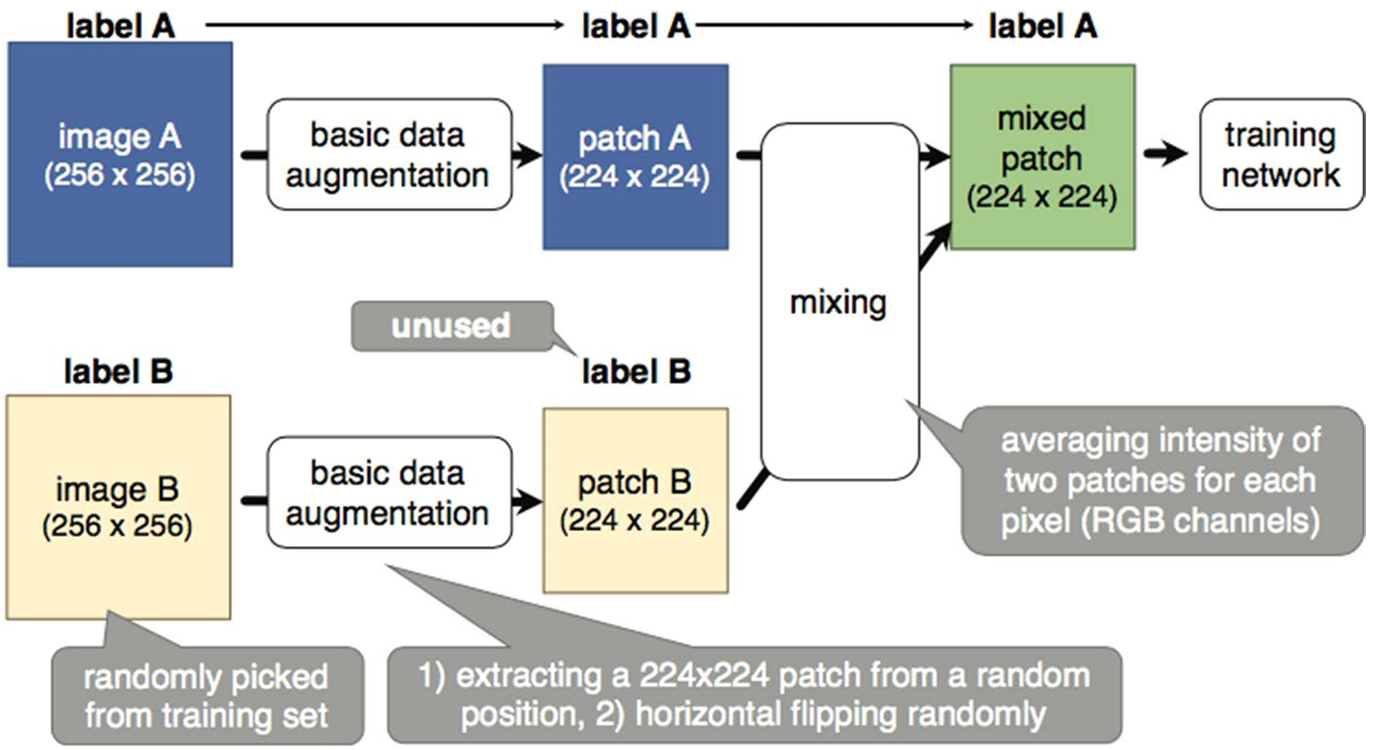
*Слика 4. Пиксели у n × n прозору се насумично премештају са вероватноћом p*

Кернел филтери су релативно недовољно истражена област у домену аугментације података. Један недостатак ове технике је тај што је веома слична унутрашњим механизмима рада CNN-ова. CNN-ови имају параметаризоване кернеле који уче оптималне начине представљања слика слој по слој. На пример, нешто попут *PatchShuffle Regularization*-а могло би се имплементирати као конволуциони слој. То се може постићи изменом параметара стандардног конволуционог слоја тако да параметри попуњавања (*padding*) очувају просторну резолуцију, а активациони слој који следи ограничава вредности пиксела у опсегу од 0 до 255. Са овим у виду, неке технике засноване на кернел филтерима могу бити боље имплементиране као део саме мреже, уместо као додатак скупу података кроз аугментацију.

### 3.2.9. Мешање слика

Мешање слика заједно просеком вредности њихових пиксела је веома неинтуитиван приступ аугментацији података. Слике добијене на овај начин неће изгледати као корисна трансформација људском посматрачу.

Две слике се насумично исецају са 256 × 256 на 224 × 224 пиксела и насумично се хоризонтално окрећу. Затим се ове слике мешају просечавањем вредности пиксела у сваком од RGB канала. То резултује мешаном сликом која се користи за тренирање модела за класификацију. Ознака која се додељује новој слици је иста као код прве насумично изабране слике (слика 5).



*Слика 5. SamplePairing стратегија аугментације*

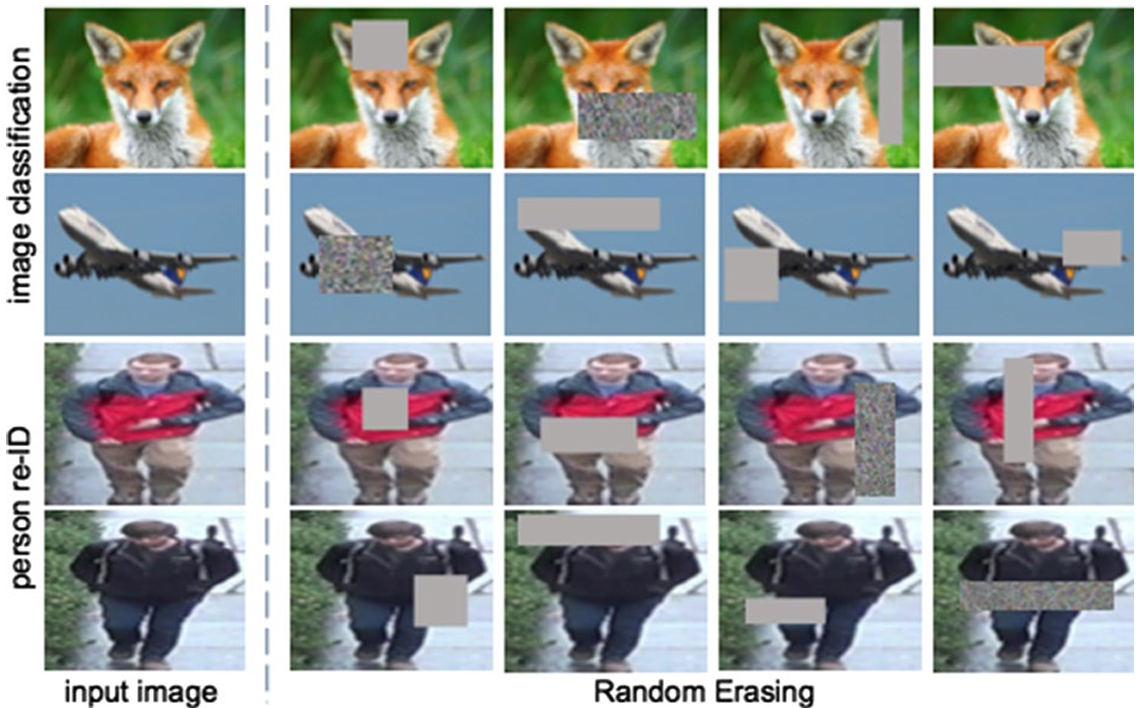
Бољи резултати су добијени када су се мешале слике из целог скупа за тренирање, уместо само из исте класе. Почевши од скупа величине *N*, *SamplePairing* производи скуп величине . Поред тога, *SamplePairing* се може комбиновати са другим техникама аугментације.

### 3.2.10. Случајно брисање (*Random Erasing*)

Случајно брисање је још једна занимљива техника аугментације података. Инспирисано механизмима регуларизације путем *dropout*-а, случајно брисање се може посматрати као еквивалент *dropout*-у, али примењено у простору улазних података, уместо да буде уграђено у архитектуру мреже. Ова техника је посебно дизајнирана да се избори са проблемима препознавања слика изазваним заклањањем (*occlusion*). Заклањање се односи на ситуације када су неки делови објекта нејасни или покривени. Случајно брисање то спречава тако што присиљава модел да научи дескриптивније карактеристике слике, чиме се избегава претренираност (*overfitting*) на одређену визуелну особину.

Поред визуелног изазова као што је заклањање, случајно брисање је обећавајућа техника која омогућава да мрежа обраћа пажњу на целокупну слику, а не само на њене делове.

Случајно брисање функционише тако што насумично изабере n × m део слике и замени га вредностима као што су све нуле (0), све 255, средња вредност пиксела, или потпуно насумичне вредности. На CIFAR-10 скупу података, ова техника је довела до смањења стопе грешке са 5.17% на 4.31%. Најефикаснији начин попуњавања избрисаних делова били су насумични пиксел вредности. Начин попуњавања и величина избрисаних делова су једини параметри који се ручно подешавају током имплементације (слика 6).

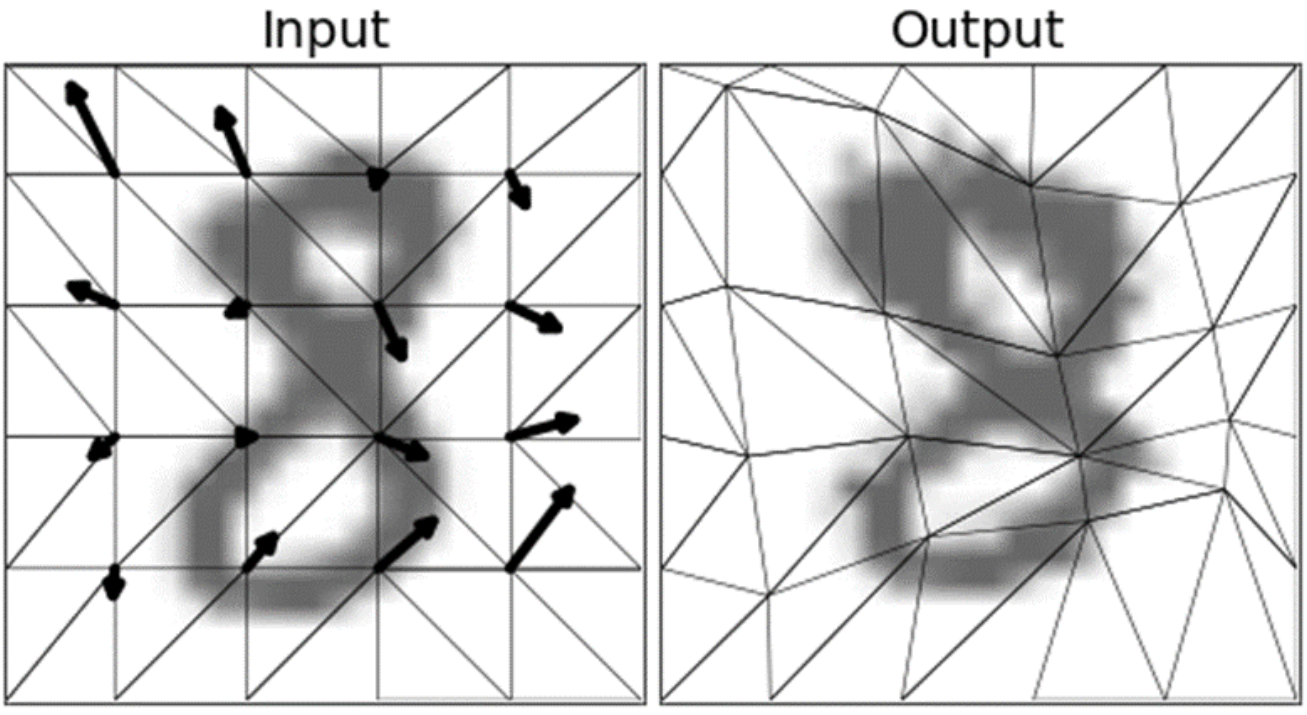


*Слика 6. Пример насумичног брисања у задацима препознавања слика*

Случајно брисање је метода аугментације података која директно спречава претренираност изменом улазног простора. Уклањањем одређених делова улаза, модел је присиљен да тражи друге дескриптивне карактеристике. Ова метода се може комбиновати са другим техникама аугментације као што су хоризонтално окретање или филтери боје. Случајно брисање је дало једну од највиших тачности на CIFAR-10 скупу.

### 3.2.11. Еластична дисторзија

Ово је примена случајних поља померања на слику, која моделује ефекат „гуме“ [5]. Код еластичног изобличења, „случајно померање“ значи да се пиксели померају унутар малог локалног окружења на слици. Математички, то се реализује помоћу поља померања које се глатко мења преко целе слике. Поље померања може се генерисати коришћењем Гаусових филтера, док параметри амплитуде и фреквенције контролишу величину насталог изобличења (слика 7).



*Слика 7. Пример еластичне дисторзије слике: (лево) оригинална слика и (десно) трансформисана*

# АУГМЕНТАЦИЈА РУКОМ ПИСАНИХ ЦИФАРА

## 4**.**1**.** Основе

Препознавање руком писаних цифара је од кључног значаја у обради слика и машинском учењу због широке примене у стварном свету, као што су обрада банковних чекова, дигитализација руком писних записа и идентификација поштанских кодова. Ефикасност и прецизност ових система у великој мери зависе од робустности коришћених модела препознавања. Због тога је унапређење ових модела важно за побољшање њихове перформансе у стварним условима.

Један од најефикаснијих начина за повећање робустности јесте аугментација података. Она обухвата креирање нових примера за обуку коришћењем различитих трансформација почетних података, као што су ротација, скалирање, транслација, обртање, исецање и еластично изобличење. Иако су ове технике широко коришћене, и даље постоји недостатак систематске процене њихове преносивости ван контролисаних скупова података као што је MNIST. Аугментација података која је инваријантна на трансформације има за циљ повећање способности модела да се генерализује са скупа за обуку на непознате податке, чиме се повећава његова прецизност и тачност.

Ово нас доводи до главног питања истраживања: „Које су најефикасније стратегије аугментације података за побољшање перформанси модела за препознавање руком писаних цифара?“ Стратегије које се користе су оне поменуте у поглављу 3. овог рада (потпоглавље 3.2).

## 4.2. Преглед трансформација над цифрама

Поменуте стратегије дају одређене резултате у процесу аугментације слика (слика 8). Свака техника се примењује са одређеним циљем и могућностима за даљу примену резултата, а најкоришћеније за овај проблем су [6]:

* Транслација

Циљ:

* Омогућава моделу да постане отпорнији на позиционирање цифре унутар слике.

Примена:

* Према дефинисаној вредности померања,
* Не комбинује се са центрирањем (због конфликта у позиционирању).
* Центрирање

Циљ:

* Уклања непотребан простор око цифре,
* Нормализује скалу цифара (све цифре добијају сличну величину),
* Убрзава процес учења.

Примена:

* Применљиво и на тренинг и на тест скуп,
* Обавезно се примењује након осталих трансформација ако се комбинује,
* Не комбинује се са транслацијом.
* Ротација

Циљ:

* Омогућава моделу да препозна цифре независно од малих нагиба (накретања током писања).

Примена:

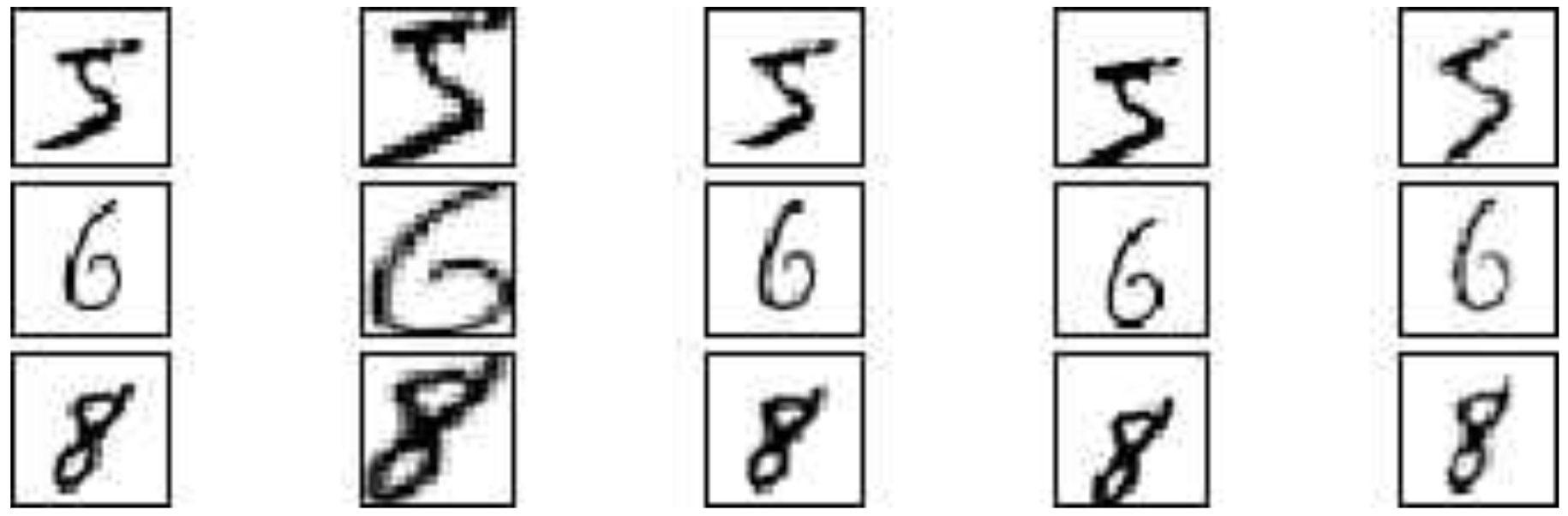
* Углови ротације су унапред дефинисани,
* Пошто се ради са пикселима (а не векторским сликама), користи се билинеарна интерполација ради добијања вредности пиксела на новим позицијама,
* Интерполација повећава варијабилност слика.
* Еластична дисторзија

Циљ:

* Омогућава моделу да буде робустан на стилске разлике у рукопису.

Примена:

* Израчунавају се матрице померања ∆x и ∆y за сваки пиксел,
* Почетне вредности су насумичне (из опсега -α до +α), а затим се „изгладе“ гаусовом конволуцијом,
* Параметар α одређује јачину деформације, а σ одређује степен глаткоће (σ ≈ 6 је оптимално),
* Завршна вредност пиксела добија се билинеарном интерполацијом.



*Слика 8. Три оригиналне MNIST слике (прва колона) и добијене слике након центрирања (друга колона), еластичне деформације (трећа колона), транслације (четврта колона) и ротације (пета колона)*

Могуће је и комбиновати трансформације, нпр. Еластична дисторзија + центрирање, Ротација + Еластична дисторзија + Центрирање, Транслација + Ротација и дупла Еластична дисторзија. Комбинације повећавају обим скупа података са 60.000 на чак 1.500.000 инстанци.

### 4.2.1. Како се шта примењује

Примене трансформација приказане су у табели 1.

| **Трансформација** | **Место у обради** | **Примена** |
| --- | --- | --- |
| Транслација | На почетку процеса | Премешта цифру унутар слике |
| Центрирање | На крају процеса | Рескалира и центрира цифру, уклања празне ивице |
| Ротација | Пре центрирања | Ротира око центра |
| Еластична дисторзија | Пре центрирање | Примењује глатку случајну деформацију преко *displacement* матрица |

*Табела 1. Приказ препорука и резултата примене техника аугментације*

## 4.3. Технологије, библиотеке и приступи

Поред основних трансформација попут ротације, скалирања и промене осветљења, постоје различите библиотеке и приступи који омогућавају истраживачима да одаберу најприкладније методе за своје поље проблема.

### 4.3.1. Универзалне библиотеке за аугментацију

Једна од најзначајнијих препорука у литератури јесте употреба универзалних библиотека које омогућавају брзо и ефикасно спровођење аугментaције [7]:

* *TensorFlow* и *Keras* - подржавају унапред дефинисане функције као што су *ImageDataGenerator*, који омогућава лако додавање трансформација током тренинга модела,
* *PyTorch* - нуди модул *torchvision.transforms* који омогућава флексибилно дефинисање ланаца трансформација,
* *Albumentations* - библиотека оптимизована за брзину и примену на великим скупoвима података; нуди напредне технике као што су *RandomCrop*, *ElasticTransform* и *GridDistortion*.

Поред наведених специјализованих библиотека за рад са сликама, могуће је комбиновати више алата ради постизања сложенијих ефеката аугментације. На пример, *scikit-learn* се често користи у припреми и организацији података - кроз технике као што су нормализација, кодирање ознака, као и *StratifiedKFold* или *train\_test\_split* функције које обезбеђују репрезентативну поделу података за тренирање и тестирање.

У комбинацији са библиотекама попут *Albumentations* или *torchvision*, *scikit-learn* омогућава лаку интеграцију аугментације у *pipeline*, док *OpenCV* (*cv2*), који је основа многих аугментационих библиотека, може бити коришћен директно за фино подешене трансформације као што су ротације, скалирање, промене боја и филтери. Ово је нарочито корисно у случајевима када је потребно ручно дефинисати или визуелизовати ефекте трансформација пре него што се аутоматизују кроз библиотеке више апстракције.

Таква комбинација - *OpenCV* за директну обраду слика, *scikit-learn* за управљање подацима и евалуацију, и напредне библиотеке као што је *Albumentations* за масовну аугментацију - омогућава изградњу флексибилних и моћних система за припрему података у задацима рачунарског вида.

### 4.3.2. Класичне технике у препознавању рукописа

У домену препознавања рукописних цифара (нпр. MNIST датасет), најчешће коришћене технике укључују [8]:

* Геометријске трансформације: ротације, транслације, скалирање и окретање,
* Фотометријске трансформације: промена контраста, осветљења, замућење и додавање шума,
* *Random Erasing*.

## 4.4. Ограничења и мане аугментације у препознавању рукописаних цифара

Иако је *data augmentation* једна од најкориснијих техника за побољшање генерализације модела, њена примена није без ограничења. У наставку су издвојени најзначајнији проблеми који се јављају у контексту препознавања рукописаних цифара.

### 4.4.1. Губитак семантичког значења

Прекомерне или неправилне трансформације могу довести до тога да узорак изгуби своје изворно значење [9]. На пример, јака ротација или агресивна еластична дисторзија може претворити цифру „6“ у нешто што више подсећа на „0“. Овакви узорци додају шум уместо корисне варијације, што доводи до повећане стопе грешке.

### 4.4.2. *Overfitting* на синтетичке податке

Иако аугментација умањује *overfitting* у односу на оригинални скуп, постоји ризик да модел постане превише везан за вештачке трансформације, а не за природну разноликост рукописа [4]. У том случају, модел учи специфичне артефакте трансформација уместо стварних образаца у цифрама.

### 4.4.3. Недовољна преносивост на реалне податке

Резултати који се добијају на контролисаним скуповима као што је MNIST често се не преносе на сложеније скупове из реалног света [10]. То значи да аугментација може побољшати тачност на експерименталним условима, али да у пракси не доноси исто повећање перформанси.

### 4.4.5. Проблем избора параметара

У процесу аугментације, избор параметара (нпр. степен ротације, ниво осветљења или интензитет еластичне дисторзије) захтева пажљиву контролу [9]. Превише агресивни параметри доводе до губитка корисних информација, док премало варијације не доноси значајне добитке у генерализацији. Ово повећава комплексност процеса тренинга и захтева бројна експериментисања.

### 4.4.6. Вештачка једноличност

Иако аугментација повећава број узорака, она не ствара праву разноликост карактеристичну за људски рукопис [4]. На тај начин може да створи илузију већег скупа података, али модел и даље може бити ограничен у сусрету са различитим стиловима писања изван тренинг скупа.

## 4.5. MNIST скуп података

MNIST база података (*Modified National Institute of Standards and Technology* база) представља велику збирку слика руком писаних цифара и често се користи за обучавање различитих система за обраду слика [11]. Такође је веома заступљена у области машинског учења, како за обучавање тако и за тестирање модела.

Настала је „ремиксовањем“ узорака из оригиналних NIST скупова података. Аутори базе су сматрали да NIST-ов скуп за тренирање, који је садржао податке запослених у америчком Заводу за статистику (*Census Bureau*), и скуп за тестирање, који је садржао податке средњошколаца из САД, нису били погодни за машинско учење због разлика у рукопису. Осим тога, оригиналне црно-беле слике из NIST-а су нормализоване да се уклопе у оквир од 28x28 пиксела и подвргнуте техници глачања ивица (*anti-aliasing*), чиме су добијени сиви тонови.

MNIST база садржи 60.000 слика за тренирање и 10.000 слика за тестирање. Половина скупа за тренирање и половина тест скупа потичу из NIST-овог оригиналног скупа за тренирање, док је преостала половина узета из оригиналног тест скупа. Аутори базе воде и листу различитих метода које су тестиране на овом скупу.

У свом оригиналном раду, користили су машину са носећим векторима (*support-vector machine*) и постигли стопу грешке од 0.8%. Проширени скуп података сличан MNIST-у, назван EMNIST, објављен је 2017. године и садржи 240.000 слика за тренирање и 40.000 за тестирање, са руком писаним цифрама и словима.

### 4.5.1. За шта се користи MNIST?

MNIST служи као полазна основа за тестирање система за обраду слика. Може се посматрати као „*hello world*“ машинског учења. Научници који се баве подацима често тренирају алгоритам на MNIST скупу једноставно како би тестирали нову архитектуру или оквир и проверили да ли исправно функционишу.

Пошто је MNIST означен скуп података, у коме су слике руком писаних цифара упарене са одговарајућим ознакама, може се користити у надгледаном учењу (*supervised learning*) за тренирање класификатора. То је добар пример, као и *ImageNet* скуп података који је креирала *Fei-Fei Li*, како добро означени скупови могу значајно унапредити развој машинског учења у ширем смислу. Још примера отворених скупова података може се пронаћи на интернету.

### 4.5.2. Како је креиран MNIST скуп података?

Проблем класификације руком писаних цифара из MNIST скупа представља стандардни скуп података који се користи у рачунарском виду и дубоком учењу.

Иако је овај скуп података у великој мери „решен“, он и даље служи као добра основа за учење и вежбање развоја, евалуације и примене конволуционих неуронских мрежа за класификацију слика од нуле. То укључује и развој поузданог система за тестирање ради процене перформанси модела, истраживање могућих побољшања модела, као и чување модела и његово касније учитавање ради предвиђања на новим подацима.

Уместо да се ослањамо на већ постојећу литературу о моделима који постижу добре резултате на овом скупу, можемо развити нови модел од почетка.

Скуп већ садржи јасно дефинисане податке за тренирање и тестирање, које можемо директно користити.

Да бисмо могли да проценимо перформансе модела у оквиру једног тренинга, можемо додатно поделити скуп за тренирање на посебне скупове за тренирање и валидацију. Перформансе модела на овим скупoвима могу се пратити током тренинга, а добијени подаци се могу графички приказати у виду кривих учења, што даје увид у то колико добро модел учи конкретан проблем.

### 4.5.3. Зашто је MNIST добар скуп података?

MNIST скуп података је изузетно добар за људе који желе да испробају технике машинског учења и методе препознавања шаблона на реалним подацима, а да при томе уложе минимално време и труд у претпроцесирање и форматирање података. Управо његова једноставност и лакоћа коришћења чине овај скуп толико широко примењивим и темељно проученим.

# ЗАКЉУЧАК

У раду је приказана и систематизована област аугментације података са нагласком на технике аугментације слика. Представљене су основне и напредне методе, уз разматрање њихових предности, ограничења и применљивости у различитим контекстима. Циљ рада је да се укаже на значај и ефикасност аугментације као кључне стратегије за побољшање перформанси модела машинског учења.

Рад је показао да добро изабране технике аугментације могу да смање потребу за огромним скуповима података и да омогуће поузданију генерализацију модела. Поред тога, указано је на важност прилагођавања аугментационих метода специфичним карактеристикама датог скупа података, као и на значај избалансиране примене техника које не изобличавају суштинске особине оригиналних примера.

Као предлог за даља истраживања издваја се анализа комбинација различитих аугментационих техника и испитивање њиховог утицаја на специфичне домене попут медицине, финансија и препознавања текста. Такође, препоручује се развој адаптивних и аутоматизованих система који би могли да сами одаберу најприкладније аугментације у складу са задатком и доступним подацима, чиме би се додатно повећала ефикасност и скратило време потребно за тренирање модела.

Додатно, потенцијал примене генеративних неуронских мрежа као што су GAN модели за синтетичко стварање квалитетних и разноврсних примера такође представља важан правац који заслужује детаљнију анализу. Овакав приступ може донети револуционарне резултате, посебно у областима где је приступ великом броју података ограничен или компликован.

Укупно, овај рад даје целовит преглед стања и перспективе развоја аугментације података, указујући на њену кључну улогу у унапређењу машинског учења и дубоког учења, као и на бројне могућности за иновативне примене у будућности.

# ЛИТЕРАТУРА

[1] Alhassan Mumunia, Fuseini Mumunib. *Data augmentation: A comprehensive*

*survey of modern approaches*. 15. новембар 2022.

[2] Jacob Murel, Eda Kavlakoglu. *What is data augmentation?* 7. мај 2024.

(<https://www.ibm.com/think/topics/data-augmentation>)

[3] aws. *What is Data Augmentation?*

(<https://aws.amazon.com/what-is/data-augmentation/>)

[4] Connor Shorten, Taghi M. Khoshgoftaar. *A survey on Image Data*

*Augmentation for Deep Learning*. 6. јул 2019.

[5] Aiym Kochkorova, Alexia Toumpa. *Data Augmentation for Handwritten Digit*

*Recognition*. 21. јануар 2025.

(<https://nhsjs.com/2025/data-augmentation-for-handwritten-digit-recognition/>)

[6] Siham Tabik, Daniel Peralta, Andrés Herrera-Poyatos, Francisco Herrera. *A*

*snapshot of image pre-processing for convolutional neural networks: case study of MNIST*. 1. јануар 2017.

[7] Adrien Payong, Shaoni Mukherjee. *Why Data Augmentation Is Key for*

*Vision, Language, and Audio AI*. 18. август 2025.

(<https://www.digitalocean.com/community/tutorials/data-augmentation-vision-language-audio-research>)

[8] Devashree. *Master Image Augmentation with 3 Widely Used Python*

*Libraries*. 21. фебруар 2024.

(<https://www.analyticsvidhya.com/blog/2022/04/master-image-augmentation-with-widely-used-python-libraries/>)

[9] Esperanza Gòmez Ramirez, Andrea Patricia Calvo Soto. *Representaciones*

*sociales de la práctica formativa en estudiantes de fisioterapia en Colombia*. август 2020.

[10] Ajian Liu, Chenxu Zhao, Zitong Yu, Jun Wan, Anyang Su, Xing Liu, Zichang

Tan, Sergio Escalera, Junliang Xing, Yanyan Liang, Guodong Guo, Zhen Lei, Stan Z. Li, Du Zhang. *Contrastive Context-Aware Learning for 3D High-Fidelity Mask Face Presentation Attack Detection*. 13. април 2021.

[11] *What is MNIST dataset?*

(<https://www.engati.ai/glossary/mnist-dataset>)

1. У тродимензионалној графици представља најмањи део тродимнзионалног простора неке сцене, који се може обрађивати или приказати (*voxel* - *volumetric pixel*) [↑](#footnote-ref-0)
2. *Generative Adversarial Network* је врста неуронске мреже у којој две мреже - генератор и дискриминатор - такмичарски уче, при чему генератор покушава да креира реалистичне податке, а дискриминатор да разликује лажне од стварних. [↑](#footnote-ref-1)
3. *Francisco Javier Moreno‑Barea* - inženjer i naučnik, doktor inženjeringa pri Univerzitetu u Malagi, specijalizovan za oblasti mašinskog učenja, veštačke inteligencije i tehnika za *data augmentation* [↑](#footnote-ref-2)