



## ОДРЕЂИВАЊЕ РАСА ПАСА И МАЧАКА УПОТРЕБОМ ДУБОКИХ НЕУРОНСКИХ МРЕЖА

### DETERMINING DOG AND CAT BREEDS USING DEEP LEARNING

Драгана Грбић, Факултет техничких наука, Нови Сад

**Област – ЕЛЕКТРОТЕХНИКА И РАЧУНАРСТВО**

**Кратак садржај** – У раду су предложена два приступа решавању проблема одређивања раса паса и мачака. Први представља класификацију у једном кораку у којем се одмах одређује раса животиње. Други представља класификацију у два корака – у првом према врсти животиње, а у другом према раси. Основу оба приступа представља употреба конволуционих неуронских мрежа, а код другог се користи и ResNet50 архитектура. Код оба приступа испитан је утицај примене активног учења.

**Кључне речи:** класификација слике, конволуционе неуронске мреже, активно учење, ResNet

**Abstract** – This paper proposes two approaches to solving the problem of determining dog and cat breeds. The first one is a one-step classification in which the breed is immediately determined. The second one represents the classification in two steps - in the first according to the species, and in the second according to the breed. Both approaches use convolutional neural networks and the second one also uses ResNet50 architecture. In both approaches, the impact of the application of active learning was examined.

**Keywords:** image classification, convolutional neural networks, active learning, ResNet

#### 1. УВОД

Људи доста времена посвећују својим кућним љубимцима, међу којима су пси и мачке једни од најпопуларнијих. Ово је потврђено великим бројем друштвених мрежа које се баве том тематиком, као што су BarkHappy [1], Dokonoko [2] и Catser [3]. Није реткост да власници паса и мачака мисле и погрешно објављују о раси свог љубимца. Зато би постојање аутоматизованог и поузданог начина да се одреди раса љубимца било од великог значаја за њих.

Проблем одређивања раса паса и мачака представља вишекласну класификацију слике, која је изазовна из више разлога.

Први лежи у чињеници да унутар једне расе паса или мачака могу постојати велике разлике по питању изгледа животиње. Некада су те разлике довољне да се уоче подрасе.

#### НАПОМЕНА:

Овај рад проистекао је из мастер рада чији ментор је био др Јелена Сливка, ванр. проф.

Други лежи у чињеници да слике паса и мачака креирају људи са различитим апаратима, умећима фотографисања и у разним ситуацијама. Стога, слике могу бити изузетно варијабилног квалитета, што по питању резолуције, осветљења и изоштрености, то и у погледу положаја и израза лица животиње на њој.

У овом раду предложено је решење проблема одређивања три расе паса и три расе мачака, које се заснива на класификацији слике употребом конволуционих неуронских мрежа и ResNet50 мреже. У циљу креирања робусног система који може решити други од проблема наведених изнад, примењена је техника аугментације података, коришћењем различитих трансформација слике. Додатно, ради смањења броја слика потребних за тренирање неуронских мрежа, а уз очување или побољшање перформанси, испитана је и примена активног учења.

У другом поглављу описан је скуп података коришћен за израду решења. У трећем поглављу описане су коришћене методе и технике. У четвртном поглављу дат је опис предложеног решења. У петом поглављу приказани су остварени резултати предложеног решења и анализирани његове грешке. У шестом поглављу набројани су најважнији закључци и наведени могући правци даљег развоја.

#### 2. СКУП ПОДАТАКА

За потребе развоја предложеног решења коришћен је The Oxford-IIIT Pet Dataset [4] који садржи укупно 7,349 слика паса и мачака. Од тога су 4,978 слике 25 раса паса, а 2,371 слике 12 раса мачака. За сваку расу присутно је по 200 слика. Свака слика у скупу аотирана је лабелом назива расе и координатама правоугаоника који уско уоквирује главу животиње.

Од укупног броја раса за потребе развоја и тестирања предложеног решења искоришћене су слике за 3 расе паса и 3 расе мачака. Одабране расе паса су Basset, Leonberger и Samoyed, а одабране расе мачака су Persian, Egyptian Mau и Siamese.

#### 3. КОРИШЋЕНЕ МЕТОДЕ И ТЕХНИКЕ

Технике машинског учења коришћене за имплементацију предложеног решења су аугментација података, активно учење и неуронске мреже. Аугментација података и активно учење искоришћени су за креирање погоднијег обучавајућег и тестног скупа, а неуронске мреже за креирање модела који ће вршити класификацију.

### 3.1. Аугментација података

Како би се добио робуснији модел, а и надоместио мали број слика по расама, над одабраним скупом података примењена је аугментација података. Аугментација искоришћена у предложеном решењу састоји се из три фазе.

У првој фази се учитавају слике и њихове анотације. Из анотације се извлаче координате правоугаоника који уско уоквирује главу животиње. Користећи те координате исеца се део који садржи главу и креира нова слика величине 224x224.

У другој фази се над новим сликама примењује низ трансформација. То су хоризонтално окретање, ротација улево и удесно до 15 степени, промена осветљености у опсегу од 0.7 до 1.4 (где вредност испод 1 представља потамњење, а вредност изнад 1 посветљење слике), исецање до 10% хоризонтално и вертикално и зумирање у опсегу од 0.95x до 1.25x.

У трећој фази се над око 30% преудонасумично одабраних трансформисаних слика примењује додатна *Gaussian Blur* трансформација. Димензије кернела којим је извршено замућење слика су 7x7.

На слици 1 приказани су примери слика добијених применом описаног поступка. Након што је из иницијалног скупа података издвојен обучавајући скуп и над њим примењена описана аугментација, добијен је коначан тренинг скуп од 6,264 слика, где за сваку расу животиње постоји по 1,044 слика.



Слика 1. Слике добијене аугментацијом података

### 3.2. Активно учење

Активно учење је техника полунадгледаног учења чији је циљ смањење података потребних за тренирање модела машинског учења, а уз очување или побољшање перформанси [5]. Суштина активног учења лежи у итеративном одабиру оних података који су најинформативнији за дати модел, односно из којих он може највише научити.

Главни проблем који се мора решити ради успешне примене ове технике јесте начин на који се за одређени податак мери његова информативност. Ради тога, може се користити низ различитих метрика, од једноставних као што је насумичан одабир, до софистициранијих решења. У овом раду испитан је утицај четири различите метрике – *BvSB*, *Bernoulli entropy* са своје две *Bavg* и *Bmax* варијанте и *EP Measure*.

Поступак технике активног учења искоришћен у овом раду састоји се из следећих корака:

1. издвајање псеудонасумично одабраног подскупа  $S$  из одговарајућег тренинг скупа  $U$ , где величина подскупа износи 20% укупног скупа
2. подела скупа  $S$  у односу 90-10 на тренинг и валидациони и тренирање модела у  $30 - 5 \times i$

епоха на њему, где  $i$  представља тренутну итерацију активног учења

3. примена модела на скуп  $U$
4. рангирање предикција употребом одговарајуће метрике и њихово сортирање по њеној вредности
5. одабир првих  $N$  слика и њихово пребацивање у скуп  $S$
6. понављање корака 2-5 у  $M$  итерација

Након што се процес активног учења заврши, крајњи тренинг скуп постаје скуп  $S$  и он се надаље користи за обучавање. Број итерација (параметар  $M$ ), број података који се лабелирају и додају у тренинг скуп у свакој итерацији (параметар  $N$ ) и конкретна метрика која се примењује представљају хиперпараметре алгоритма и њихове вредности одређене су емпиријски.

### 3.3. Неуронске мреже

За имплементацију модела који врше класификацију животиња употребљена је конволуциона неуронска мрежа са самостално развијеном архитектуром и *ResNet50* архитектура. Конволуциона мрежа је тренирана у 30 епоха, а *ResNet50* мрежа је претренирана на *ImageNet 1000* [6] скупу података и није накнадно тренирана на скупу коришћеном у раду.

Самостално развијена мрежа састоји се из 6 конволуционих слојева, праћених једним потпуно повезаним слојем. Конволуциони слојеви користе *ReLU* активациону функцију. Након сваког конволуционог слоја следи по један *Max Pooling* сажимајући слој и један *Dropout* слој. У последњем слоју коришћена је *Softmax* активациона функција.

## 4. ОПИС ПРЕДЛОЖЕНОГ РЕШЕЊА

Током развоја предложеног решења испробана су два приступа класификацији раса паса и мачака. Следи опис оба приступа.

### 4.1. Први приступ – класификација у једном кораку

Идеја првог приступа је одређивање раса паса и мачака у једном кораку, односно без претходног закључивања о томе да ли се на слици налази пас или мачка. У ту сврху коришћена је самостално развијена конволуциона неуронска мрежа. Мрежа је обучена да препознаје 6 класа – 3 расе паса и 3 расе мачака. У наставку рада мрежа ће се означавати са CNN-1.

Предност овог приступа је чињеница да је за његову имплементацију потребна само једна неуронска мрежа. Мана је то што мрежа може направити јако грубу грешку тиме што слику мачке може класификовати као слику пса и обрнуто.

### 4.2. Други приступ – класификација у два корака

Други приступ врши класификацију слика паса и мачака у два корака. У првом кораку врши се класификација по принципу пас-мачка. У ту сврху коришћена је претренирана *ResNet50* неуронска мрежа. У другом кораку се, на основу резултата првог корака, слика прослеђује моделу који је

специјализован да обавља класификацију само по врстама паса или мачака. У ту сврху коришћене су две самостално развијене конволуционе неуронске мреже. Једна врши класификацију само паса и у даљем тексту означаваће се са CNN-D. Друга врши класификацију само мачака и у даљем тексту означаваће се са CNN-C. Обе мреже трениране су над подкуповима укупног тренинг скупа, који садрже само слике паса, односно мачака.

Како *ResNet50* неуронска мрежа не мора препознати ни пса ни мачку на слици, због чињенице да је тренирана да препознаје и друге објекте, у таквим случајевима се слика у другом кораку прослеђује и CNN-D класификатору и CNN-C класификатору. Као крајње решење класификације узима се одлука оног класификатора који је у своју одлуку био сигурнији, односно чија је предикција имала већу вероватноћу припадања предиктованој класи. Разлог за ово јесте интуиција да ће класификатор мачака бити мање сигуран у своју одлуку од класификатора паса уколико му се проследи слика пса и обрнуто.

### 4.3. Испитивање метрика активног учења

Код оба приступа испитан је утицај активног учења приликом обучавања модела. Испитивање је извршено за сваки од модела који користе самостално развијену архитектуру (CNN-1, CNN-D и CNN-C). Како се *ResNet50* неуронска мрежа из другог приступа користи претренирана, односно не обучава додатно на скупу коришћеном у раду, за њу није вршено испитивање.

Вредност параметра  $M$ , који представља број итерација активног учења, за сваки модел је одабран да буде 4. Вредност параметра  $N$ , који представља број података који се у свакој итерацији додају у тренинг скуп, одређен је за сваки модел и итерацију активног учења понаособ (табела 1).

Табела 1. Вредност параметра  $N$  по итерацијама и моделима

Итерација	CNN-1	CNN-D	CNN-C
1	1,050	525	525
2	1,050	525	525
3	900	450	450
4	900	450	450

За евалуацију најбоље метрике информативности коришћена је тачност и *loss* на валидационом скупу за другу итерацију активног учења. Разлог за одабир ове итерације лежи у чињеници да су сви модели већ на њој остварили јако добре перформансе, а уз коришћење само око 53% укупног тренинг скупа. Метрике које су се најбоље показале по моделима су:

- CNN-1 – *Bmax*, 99.6% тачност, 0.0145 *loss*
- CNN-D – *EP Measure*, 100% тачност, 0.0029 *loss*
- CNN-C – *BvSB*, 99.8% тачност, 0.0136 *loss*

## 5. РЕЗУЛТАТИ И ДИСКУСИЈА

Оба приступа евалуирана су над тестним скупом користећи тачност као метрику. Следи опис поступка евалуације.

### 5.1. Тестни скуп

Тестирање перформанси предложеног решења извршено је над тестним скупом који се састоји од 420 слика. Скуп је формиран тако што је из иницијалног скупа података издвојено по 7 слика за сваку од 6 раса животиња и формиран почетни скуп од 42 слике. Након тога је применом истих трансформација коришћених за аугментацију тренинг скупа добијен крајњи скуп од 420 слика.

Разлог за употребу оваквог тестног скупа јесте жеља за тестирањем робусности модела, односно отпорности на деформисаност слика, слика ниског квалитета, разлике у нивоу осветљености и слично.

### 5.2. Поступак тестирања и остварени резултати

Упоредо су тестирана оба приступа, са и без примене активног учења. Модели тренирани без примене активног учења обучавањем су над целим тренинг скупом. Модели тренирани уз коришћење активног учења обучавањем су уз примену две итерације ове технике и метрике која се за њих показала најбољом.

Резултати добијени приликом тестирања решења приказани су у табелама 2 и 3. Обе табеле садрже добијене вредности тачности за сваки приступ и модел. Табела 2 садржи резултате у случају када није примењено активно учење, а табела 3 у случају када јесте. Метрика означена са RN-2S представља коначну тачност другог приступа, односно тачност која се добије када се цео систем евалуира, а не само појединачне мреже.

Табела 2. Резултати модела и приступа без примене активног учења

	CNN-1	CNN-D	CNN-C	ResNet	RN-2S
Тачност (%)	93,33	97,62	97,62	96,43	95,00

Табела 3. Резултати модела и приступа у случају примене активног учења

	CNN-1	CNN-D	CNN-C	ResNet	RN-2S
Тачност (%)	97,86	96,67	95,24	96,43	95,24

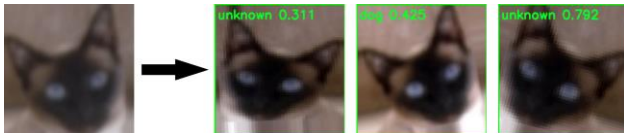
На основу добијених резултата може се закључити да се други приступ показао бољим, у случају да се не користи активно учење. Додатно, може се уочити да су модели добијени применом активног учења дали упоредиве, а у случају првог приступа и осетно боље перформансе у поређењу са моделима добијеним без примене ове технике.

Иако су оба приступа дали упоредиве резултате, предност у погледу даљег развоја требало би дати другом приступу, односно класификацији у два корака. Разлог за то лежи у *ImageNet 1000* скупу података на коме је претренирана *ResNet50* неуронска мрежа. Наиме, овај скуп је изузетно разноврстан по питању класа које садржи, те би у случају додавања нове врсте животиње постојала велика вероватноћа да за њу постоји одговарајућа класа у *ImageNet 1000* скупу. Самим тим, не би било потребно дотренирати коришћену мрежу, већ би само било довољно направити модел који ће вршити класификацију по расама у оквиру нове врсте.

### 5.3. Анализа грешака модела

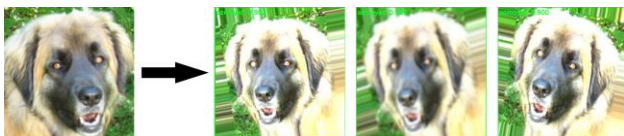
Грешке које су модели начинили у највећем броју случајева лако су објашњиве и могу се приписати деформисаности слика у тестном скупу, што због броја примењених трансформација и њиховог интензитета, што због ниског квалитета оригиналне слике.

Тако је *ResNet50* модел погрешно класификовао 9 од 10 слика које су добијене од једне исте полазне која је малих димензија, лошег осветљења и велике замућености (слика 2). Од тих 9 грешака, 7 је начињено јер модел није успео на слици препознати ни пса ни мачку, док је у једном случају препознао пса уместо мачке.



Слика 2. Грешке модела (десно) услед ниског квалитета полазне слике (лево)

Додатно, модел CNN-D добијен без примене активног учења је у случају слике са јаким осветљењем, а од које су трансформацијама добијене додатно посветљене слике, погрешно класификовао пса расе *Leonberger* као пса расе *Samoyed* (слика 3). Ово се може приписати јакој деформацији слике услед повећања осветљености, те самим тим повећања и беле боје на псу, што је навело модел да донесе погрешну одлуку.



Слика 3. Грешке модела на посветљеним сликама (десно) добијеним од јако осветљене полазне слике (лево)

### 6. ЗАКЉУЧАК

Размотрен је проблем одређивања раса паса и мачака и зашто је он сам по себи важан. Предложено је решење са два различита приступа која се заснивају на употреби конволуционих неуронских мрежа. Описан је коришћен скуп података, начин његове аугментације, концепт активног учења, као и коришћене неуронске мреже. Предложено решење је тестирано и резултати верификовани.

На основу добијених резултата може се закључити да се други приступ предложеног решења показао бољим. Додатно, упоредном анализом резултата добијених са и без примене активног учења закључује се да је примена ове технике оправдана, будући да су постигнути упоредиви или бољи резултати, а уз знатно смањење тренинг скупа.

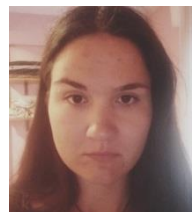
Дакле, класификација у два корака, где се у првом одређује врста, а у другом раса пса или мачке, у комбинацији са техником активног учења и примене аугментације података, представља добро решење проблема одређивања раса паса и мачака.

Једно од могућих проширења предложеног решења могла би бити употреба напреднијих архитектура неуронских мрежа, као и проширење по питању подржаних врста и раса животиња. Додатно, решење би се могло проширити додавањем подршке за детекцију животиња на слици, као и анализом грешака модела уз помоћ различитих техника визуелизације модела.

### 7. ЛИТЕРАТУРА

- [1] <https://barkhappy.com> (приступљено у марту 2022.)
- [2] <https://www.dokonoko.jp/en> (приступљено у марту 2022.)
- [3] <https://www.catster.com> (приступљено у марту 2022.)
- [4] Zisserman, Andrea Vedaldi Andrew, and C. V. Jawahar. "Cats and dogs." IEEE conference on computer vision and pattern recognition. 2012.
- [5] Cohn, David, Les Atlas, and Richard Ladner. "Improving generalization with active learning." Machine learning 15.2 (1994): 201-221.
- [6] Deng, Jia, et al. "Imagenet: A large-scale hierarchical image database." 2009 IEEE conference on computer vision and pattern recognition. Ieee, 2009.

### Кратка биографија:



Драгана Грбић рођена је у Новом Саду 1998. године. Дипломирала је на Факултету техничких наука из области Софтверско инжењерство и информационе технологије 2021. године. Мастер рад из исте области одбранила је 2022. године.