

# Препознавање писаних јапанских слова употребом конволуционих неуронских мрежа

Марија Ћурчић

R2 28/2021

marija.curcic07@gmail.com

**Анстракт** — Јапанска слова писана на традиционални калиграфски начин доста се разликују од слова писаних на савремени начин, на који је већина људи навикло, што чини да их људи тешко препознају. У овом раду је представљено решење које обавља детекцију писаних јапанских слова са слике и њихову класификацију на одговарајућа слова у регуларном облику. Решење се састоји из три модула: модул за обраду података, модул за детекцију и модул за класификацију. У модулу за обраду података примењене су технике претпроцесирања слика. За детекцију слова коришћена је R-CNN архитектура са *AlexNet* конволуционом неуронском мрежом и постигнута је тачност од 99% на обучавајућем скупу и 98% на тестном скупу. За класификацију је коришћена конволуциона неуронска мрежа која је описана у раду, и она достиже тачност од 99% на обучавајућем и 94% на тестном скупу.

**Кључне речи** — детекција и класификација писаних слова; *kuzushiji*; CNN; R-CNN; *AlexNet* рачинарска визија

## I. Увод

Кузушиђи (јап. 崩し字 – *kuzushiji*: „деформисани симболи“) је назив за курзивно јапанско писмо, односно ручно писана јапанска слова [1]. Овим писмом су писани многи историјски документи у Јапану, и оно се визуелно доста разликује од писма које се користи данас. С обзиром на то да велик број Јапанаца не уме да чита кузушиђи, системи за превођење оваквих докумената на савремено јапанско писмо су од великог значаја.

Циљ овог рада је детекција и класификација јапанских слова и симбола са слика страница које садрже текстове писане курзивним јапанским писмом, коришћењем конволуционих неуронских мрежа.

У поглављу II представљени су радови који су се бавили истим или сличним проблемима и њихови приступи и резултати. У поглављу III описан је скуп података који је коришћен за решавање овог проблема. У поглављу IV описана је структура решења, технике које су коришћене за претпроцесирање и обраду података, и методе детекције и класификације. У поглављу V представљени су резултати. Поглавље VI је закључак и сумаризација рада.

## II. СРОДНИ РАДОВИ

Рад [2] бави се детекцијом писаних јапанских слова коришћењем различитих методе детекције објеката. Прва метода детекције је FDGM (*Frame Group Decision Method*)

која је заснована на YOLOv3 систему. Друга метода је FDMBSS (*Frame Decision Method Based on Selective Search*). Скуп података који је коришћен у овом раду сачињен је комбиновањем CODH скупова података и података са PRMU конкурса. Овај скуп садржи 77000 слика са три различита слова која се узастопно понављају. Дефинисана су два критеријума за евалуацију решења: Y-CR, који представља разлику граничних оквира у вертикалном правцу, и LER (*Label Error Rate*). Модели постижу ниске вредности за LER, али недовољан Y-CR, док FDMBSS постиже најбоље резултате за Y-CR, али је LER превелик.

У раду [3] описана је методологија препознавања курзивних јапанских слова користећи приступ заснован на вишеструким карактеристикама и моделе дубоког учења. Коришћен је CODH (*Center for Open Data in the Humanities*) скуп података који садржи преко 600,000 слика и 3929 класа за различита слова. Процес сегментације података састоји се из следећих корака: претпроцесирање, детекције ивица *Canny* алгоритмом, техника сегментације контура, примена *convex hull* алгоритма и *threshold* методе. Затим је одрађено издвајање три врсте својстава: *zonal features* (ZF), *structural features* (SF) и *invariant moments* (IM). За класификацију су испробане две методе: SVM, са којим је постигнута тачност од 87,4% и класична неуронска мрежа, која постиже тачност од 90%.

## III. СКУП ПОДАТАКА

Скуп података који је коришћен за детекцију и класификацију преузет је са адресе [4]. Овај скуп података је лабелиран и подељен на скупове за обучавање и тестирање. Садржи 3605 слика за обучавање и 1730 слика за тестирање. На свакој слици је папирна страница која садржи неки текст написан курзивним јапанским словима, а неке странице могу да садрже и илустрације (Слика 1).



Слика 1 – примери слика из скупа података

За слике из тестног и обучавајућег скупа постоје CSV фајлови са лабелама за сваку слику, које садрже назив слике, координате, висину, ширину и ознаку *Unicode* карактера за свако слово са слике.

#### IV. МЕТОДОЛОГИЈА

У овом поглављу описане су технике које су коришћене за претпроцесирање података, као и методологија и архитектура решења.

##### 1. Претпроцесирање

Пре употребе скупа података примењени су одређени кораци претпроцесирања.

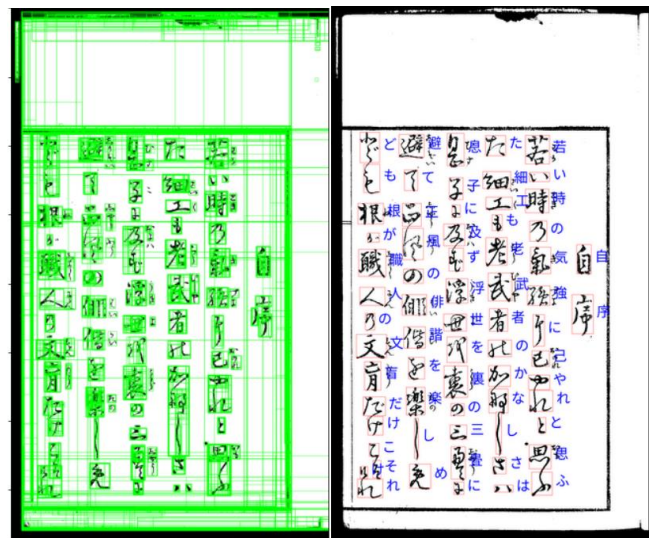
За обраду слике коришћена је OpenCV библиотека. Свака слика је прво конвертована из BGR у RGB формат. Висина и ширина су промењене тако да све слике буду истих димензија. Приликом промене димензија слика, измењене су вредности за граничне оквири у фајлу са лабелама тако да одговарају новим димензијама слика. Затим је одрађено конвертовање слике у *grayscale*, односно нијансе сивих. На сваку слику је примењен *threshold*, коришћењем *Otsu* методе за проналажење прага. На Слика 2 приказане су слике из скупа података које су прошле кроз претпроцесирање.



Слика 2 – примери слика након претпроцесирања

##### 2. Детекција

Модул за детекцију има улогу да пронађе граничне оквири за слова и симболе који се налазе на слици. За детекцију је коришћена R-CNN (*Region-based Convolutional Neural Network*) архитектура са *AlexNet* конволуционом неуронском мрежом [5]. Сегментација је одрађена употребом метода из OpenCV библиотеке. Коришћена је IoU (*Intersection over Union*) метрика, чије су вредности рачунате на основу података из фајла са лабелама. Гранични оквири са IoU од 0.7 или више су сматрани прихватљивим. На левој слици је приказана слика са граничним оквирима добијеним сегментацијом, а на десној слици је визуелни приказ граничних оквира и одговарајућих слова из фајла са лабелама (Слика 3).

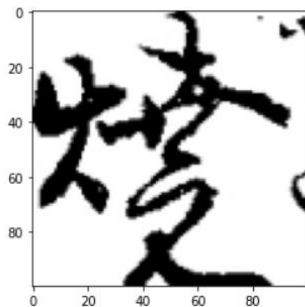


Слика 3 – Гранични оквири

##### 3. Класификација

Задатак модула за класификацију је да свакој слици на улазу додели једну класу која представља неко јапанско слово. Из скупа података који је добијен обрадом слика

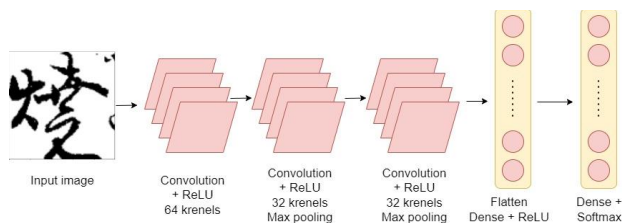
описаном у делу за претпроцесирање, издвајају се слике појединачних слова исецањем слика страница на основу података за граничне оквире и ознаке за свако слово у фајлу са лабелама. Све исечене слике се пребацују на величину од 100×100 (Слика 4).



Слика 4 – Улазна слика за класификатор

У скупу података који је представљен у поглављу постоји 1400 различитих слова и симбола. Због превелике количине података, времена које је потребно за обучавање неуронске мреже и неизбалансираности података по класама, скуп који се користи у овом модулу је редукован на мање слика и мањи број класа. Од првих 100000 исечених слика са појединачним словима издвојене су само слике слова која се појављују 300 или више пута. На тај начин је добијено скоро 64500 слика и 59 различитих класа.

Као класификатор коришћена је конволуциона неуронска мрежа чија је архитектура приказана на Слика 5.



Слика 5 – Структура конволуционе неуронске мреже за класификацију слова

Први слој је конволуциони слој са 64 кернела димензија 3×3. После њега следе два блока који се састоје од једног конволуционог слоја са 32 кернела, сажимајућег слоја који примењује *max pooling* технику са кернелом димензије 2×2, и *dropout* регуларизације од 0.4. Сви конволуциони слојеви примењују ReLU активациону функцију. Између ових слојева је примењивана *Batch* нормализација. Затим следи слој за поравнање и потпуно повезани слој са ReLU активационом функцијом. Излазни слој је *softmax* са 59 класа.

## V. РЕЗУЛТАТИ

Модел за детекцију обучаван је у 10 епоха. Скуп података је подељен на обучавајући, валидациони и тестни скуп тако да обучавајући чини 80% података, а валидациони и тестни по 10%.

Модел за класификацију обучаван је у 20 епоха. Скуп података је подељен на обучавајући, валидациони и тестни скуп, где обучавајући скуп чини 70% података, валидациони 20% и тестни 10%.

За евалуацију модела током обучавања и тестирања коришћена је метрика *accuracy*. Резултати за обучавање и тестирање за оба модула приказани су у Табела 1.

	Обучавање	Тестирање
<b>R-CNN</b>	0,99	0,98
<b>CNN класификатор</b>	0,99	0,94

Табела 1 - Резултати

## VI. ЗАКЉУЧАК

У овом раду су представљене методе детекције и класификације писаних јапанских слова. Решење је подељено на модул за обраду слика, модул за детекцију и модул за класификацију слова и знакова. Модели су обучавани и евалуирани на једном делу скупа података који је описан у раду и дали су задовољавајуће резултате.

Ово решење се може применити у транскрипцији старих ручно писаних јапанских докумената у текстове на савременом јапанском писму. У овом раду класификатор је обучен тако да препознаје 59 симбола, али скуп података који је представљен је довољно обиман да се ово решење прошири на већи број класа.

## Литература

- [1] Introduction to kuzushiji 崩し字 : <http://naruhodo.weebly.com/blog/introduction-to-kuzushiji>
- [2] Y. Tang, K. Hatano, E. Takimoto, "Recognition of Japanese Historical Hand-Written Characters Based on Object Detection Methods", 2019.
- [3] Aravinda C.V, L. Meng, A. Masahiko, Udaya Kumar Reddy K.R, A. Prabhu, "A Complete Methodology for Kuzushiji Historical Character Recognition using Multiple Features Approach and Deep Learning Model", 202.
- [4] Kuzushiji Recognition: <https://www.kaggle.com/competitions/kuzushiji-recognition/data>
- [5] Aston Zhang, Zachary C. Lipton, Mu Li, and Alexander J. Smola: "Dive into Deep Learning", 2021