SVEUČILIŠTE U ZAGREBU
**FAKULTET ELEKTROTEHNIKE I RAČUNARSTVA**

ZAVRŠNI RAD br. 1652

# POBOLJŠANJE REZOLUCIJE SLIKA TEKSTA

Jura Hostić

Zagreb, lipanj 2024.

SVEUČILIŠTE U ZAGREBU
**FAKULTET ELEKTROTEHNIKE I RAČUNARSTVA**

ZAVRŠNI RAD br. 1652

# POBOLJŠANJE REZOLUCIJE SLIKA TEKSTA

Jura Hostić

Zagreb, lipanj 2024.

**SVEUČILIŠTE U ZAGREBU**
**FAKULTET ELEKTROTEHNIKE I RAČUNARSTVA**

Zagreb, 4. ožujka 2024.

# ZAVRŠNI ZADATAK br. 1652

| | |
|---|---|
| Pristupnik: | **Jura Hostić (0036538357)** |
| Studij: | Elektrotehnika i informacijska tehnologija i Računarstvo |
| Modul: | Računarstvo |
| Mentor: | prof. dr. sc. Siniša Šegvić |

Zadatak: **Poboljšanje rezolucije slika teksta**

Opis zadatka:

Pobošavanje rezolucije zanimljiv je problem računalnog vida s mnogim uzbudljivim primjenama. Zbog važnosti natpisa u našem svakodnevnom životu, posebno su zanimljive metode specijalizirane za slike teksta. Ovaj rad istražit će primjenjivost postojećih metoda na različitim zadatcima. U okviru rada, potrebno je odabrati okvir za automatsku diferencijaciju te upoznati biblioteke za rukovanje matricama i slikama. Proučiti i ukratko opisati postojeće duboke arhitekture utemeljene na konvolucijama i pažnji. Vrednovati generalizacijsku moć postupaka iz literature na javno dostupnim i privatnim slikama. Procijeniti složenost učenja modela. Prikazati i ocijeniti provedene eksperimente. Radu priložiti izvorni i izvršni kod razvijenih postupaka, ispitne slijedove i rezultate, uz potrebna objašnjenja i dokumentaciju. Citirati korištenu literaturu i navesti dobivenu pomoć.

Rok za predaju rada: 14. lipnja 2024.

# Sadržaj

# 1. Uvod

U digitalno doba, prepoznavanje teksta na slikama je važan zadatak u svakodnevnom životu. Ima mnoge primjene poput pretraživanja dokumenta, digitalizacija fizičkih papira ili čitanje znakova pri autonomnoj vožnji. Kvaliteta tih slika igra ključnu ulogu u točnosti prepoznavanja, stoga slike niže rezolucije često predstavljaju probleme zbog nedostatka detalja na slikama. Iz tog razloga, metode za poboljšanje rezolucije slika koje čine tekst lakše čitljivim su ključne za daljnji razvoj tehnologije.

U ovom radu su istraživane i uspoređene različite metode poboljšanja rezolucije slika s fokusom na njihovu primjenu u svrhu jednostavnijeg čitanja teksta. Posebna je pažnja obraćena na metode temeljene na dubokom učenju. Analizirano je nekoliko novijih modela dubokog učenja iz literature, što uključuje modele specifično dizajnirane za tekst te općenitiji model za poboljšanje rezolucije. Cilj je utvrditi koji pristupi postižu najbolje rezultate u poboljšanju čitljivosti teksta na slikama.

Rad je organiziran u 6 poglavlja nakon kojih se nalazi popis literature i sažetak. Drugo poglavlje daje pregled osnovnih koncepata računalnog vida te korištene tehnologije u sklopu rada. Treće poglavlje opisuje korištene metode, skupove podataka i metodologiju. U četvrtom poglavlju je opisana organizacija eksperimenata u svrhu reprodukcije rezultata. Peto poglavlje sadrži rezultate svih eksperimenata nad podskupovima podataka te njihovu analizu i usporedbu. Na kraju, u petom poglavlju se donose zaključci i raspravlja o budućnosti područja.

# 2. Povećanje rezolucije teksta

## 2.1. Računalni vid

Računalni vid je interdisciplinarno područje koje se bavi razvojem algoritama i tehnika koje omogućavaju računalima da razumiju svijet oko sebe putem vizualnih podražaja nalik ljudskom vidu. Ovo područje obuhvaća razne tehnike, poput obrade slika, manipulacije, analize te izvlačenje informacija iz vizualnih podataka. Upotrebe su raznolike, omogućuje prepoznavanje oblika, označavanje bolesti te između ostaloga i povećanja rezolucija slika, što je tema kojom se bavimo u ovom radu.

Računalni vid ima ključnu ulogu u povećanju rezolucije teksta. Tradicionalne tehnike poboljšanja slika koristile su algoritme za postizanje cilja [13], ali razvojem tehnologije razvili su se novi načini zasnovani na dubokom učenju. Kompleksnost i varijabilnost teksta predstavlja problem za tradicionalne algoritme. Duboko učenje zato, s mogućnošću prepoznavanja kompleksnih obrazaca je idealno za takav problem.

Duboko učenje, posebice konvolucijske neuronske mreže (CNN), revolucioniralo je područje računalnog vida. CNN-ovi su se pokazali iznimno uspješnima u učenju i prepoznavanju složenijih značajki sa slika poput oblika, tekstura ili teksta.

U ovome radu, istražene su postojeće metode koje su se bavile povećanjem rezolucije slika kako bi poboljšali mogućnosti prepoznavanja teksta.

## 2.2. Alati za implementaciju računalnog vida

Za implementaciju algoritama računalnog vide te dubokog učenja koriste se posebni alati i okviri. Sve implementacije koje će se proučavati u ovome radu su napisane u Pythonu s pomoću dodatnih alata i okvira za njega.

### 2.2.1. Python [1]

Python je programski jezik koji se ističe svojom jednostavnošću korištenja te bogatom kolekcijom biblioteka. S pomoću raznih biblioteka i okvira za Python, on je idealan za primjenu u strojnom učenju i računalnom vidu, zbog čega je postao standardni alat u tim slučajevima.

### 2.2.2. PyTorch [2]

PyTorch je jedan od najkorištenijih okvira za duboko učenje. On omogućava fleksibilno i jednostavno definiranje neuronskih mreža te lagano učenje i evaluaciju samih. Zbog njegove podrške za automatsku diferencijaciju, omogućuje učinkovitu implementaciju algoritama za učenje, znatno bolju nego što bi se moglo samo u Pythonu. Kako je vrlo raširen, ima vrlo dobro razvijene dodatne biblioteke koje omogućavaju izradu rješenja za razne upotrebe, što uključuje i računalni vid. Jedna od bitnih značajki je da omogućuje korištenje grafičke kartice za paralelno računanje što znatno ubrzava proces dubokog učenja. Sve implementacije koje ćemo proučavati koriste PyTorch, ali to nije jedini okvir za duboko učenje za Python, također je vrlo popularan i TensorFlow.

### 2.2.3. NumPy [3]

NumPy je biblioteka za matematičko računanje koja omogućuje brzo i jednostavno računanje s višedimenzionalnim nizovima (tenzorima). Zbog toga je jedan od osnovnih alata pri dubokom učenju. U nastavku su dana dva koda, jedan od njih napisan s pomoću NumPy-a, drugi bez njega. Oba koda služe istoj svrsi, ali imaju značajno drugačiju brzinu izvedbe kao što ćemo vidjeti u nastavku.

Kod u Pythonu koji kreira dvije matrice N x N s nasumičnim vrijednostima te ih zatim pomnoži:

```
1  def create_and_multiply_matrices_with_vanilla_python (dimensions: int)
   -> list[list[float]]:
2    A = [[random.random() for _ in range(dimensions)] for _ in
   range(dimensions)]
3    B = [[random.random() for _ in range(dimensions)] for _ in
   range(dimensions)]
```

```
4
5       C = [[0 for _ in range(dimensions)] for _ in range(dimensions)]
6       for i in range(dimensions):
7         for j in range(dimensions):
8           for k in range(dimensions):
9             C[i][j]+=A[i][k]*B[k][j]
10
11      return C
```

Isti kod koristeći NumPy:

```
1     def create_and_multiply_matrices_with_numpy(dimensions: int) ->
      np.ndarray:
2     A = np.random.rand(dimensions, dimensions)
3     B = np.random.rand(dimensions, dimensions)
4     C = np.dot(A, B)
5
6     return C
```

Razlika u brzini izvršavanju te dvije funkcije se vidi na slici 2.1. Obje funkcije su pozvane s argumentom N = 1000, te se očituje ogromna razlika (red veličine 103) u brzini izvođenja što pokazuje zašto je NumPy neprocjenjiv pri dubokom učenju.

```
Numpy:
Time taken: 0.033235073089599 61 seconds
Vanilla Python:
Time taken: 51.937084913253784 seconds
```

**Slika 2.1.** Ispis vremena izvršavanja navedenih funkcija s argumentom N=1000

### 2.2.4. CUDA [4]

CUDA je platforma za paralelno računanje razvijena od strane NVIDIA-e. Ona omogućava brzo izvršavanje koda korištenjem velikog broja jezgri grafičkih kartica za paralelno računanje. CUDA se koristi u sklopu PyTorch-a kako bi se iskoristile mogućnosti grafič-

kih kartica. Eksperimenti su pokretani s pomoću NVIDIA GTX 1060 grafičke kartice lokalno te NVIDIA A100 GPU unutar Google Colaba.

### 2.2.5. Jupyter Notebook [5] i Google [6]

Jupyter Notebook je interaktivno razvojno okruženje koje omogućuje lagano dijeljenje i pokretanje Python koda namijenjenog za razne uporabe. Unutar njega će biti pokretani neki od eksperimenata, zbog njegove modularnosti koja omogućuje lagano iteriranje. Google Colab je platforma koja omogućuje pokretanje Jupyter bilježnica na oblaku. Ona omogućuje pristup znatno bržim grafičkim karticama što znatno ubrzava proces učenja i validacije. Unutar nje će biti pokretani zahtjevniji eksperimenti koji nisu mogući lokalno zbog nedostatka resursa.

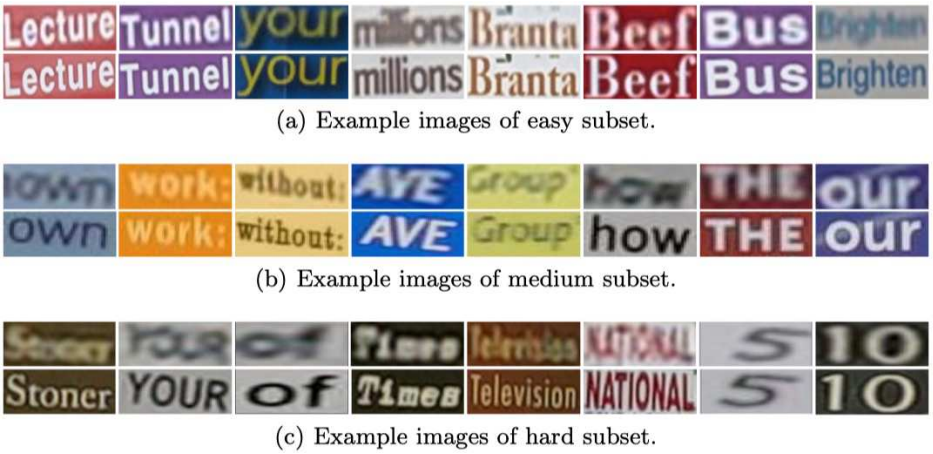
# 3. Korištene implementacije i skupovi podataka

Unutar ovog rada će se usporediti uspješnost različitih postojećih implementacija poboljšanja rezolucije. Implementacije će povećati rezoluciju slika iz TextZoom skupa podataka. CRNN model će zatim pročitati tekst s tih slika i usporediti ga s točnim tekstom. Kao skup podataka izabran je TextZoom skup podataka zbog svoje realistične reprezentacije slika niske rezolucije.

## 3.1. TextZoom [7]

Kako bi se procijenila učinkovitost metoda u odnosu na stvarno stanje, potreban je skup podataka koji sadrži slike niske rezolucije (LR) te visoke rezolucije (HR). Postoje razni umjetno kreirani skupovi gdje su LR slike generirane iz HR slika algoritamski. Problem takvih skupova podataka je što LR slike ne odražavaju realne uvjete snimanja, što može dovesti do nerealno dobrih rezultata prilikom procjene točnosti.

Iz tog razloga je izabran TextZoom (slika 3.1.). Unutar tog skupa, svaka slika je fotografirana dva puta s različitim žarišnim duljinama kako bi se postigli parovi slika niže i više kvalitete. Za potrebe ovog rada, korištene su isključivo slike iz "test" podskupa. Zbog načina rada CRNN modela, skup je dodatno filtriran kako bi sadržavao isključivo slike teksta koji se sastoji od slova, brojeva te znaka "-".




(a) Example images of easy subset.

(b) Example images of medium subset.

(c) Example images of hard subset.

**Slika 3.1.** Primjer podataka iz TextZoom seta podataka[14]

## 3.2. Vlastite slike

Za proširenje eksperimenta te testiranje na dodatnim stvarnim slikama kreiran je vlastiti skup podataka. Vlastiti skup podataka se sastoji od 12 slika niske kvalitete te 12 slika visoke kvalitete. Slike su snimljene uređajem iPhone 14 Pro Max s različitim postavkama fokusa, pri čemu su slike visoke kvalitete snimljene u fokusu, a slike niske kvalitete blago izvan fokusa. Iz ukupno 10 snimljenih fotografija (pet u fokusu, pet izvan fokusa) odabrano je 12 isječaka teksta za skup podataka. Skup podataka sadrži raznovrsne slike, uključujući slike samo sa slovima (npr. 3.2. i 3.3.), samo brojevima, kombinacijom slova i brojeva te iznimno zahtjevne slučajeve poput 3.4. i 3.5.


**Slika 3.2.** Slika iz vlastitog skupa niske kvalitete


**Slika 3.3.** Slika iz vlastitog skupa visoke kvalitete

**Slika 3.4.** Zahtjevna slika niske kvalitete iz vlastitog skupa


**Slika 3.5.** Zahtjevna slika visoke kvalitete iz vlastitog skupa

## 3.3. Convolutional Recurrent Neural Network [8]

Nakon povećanja rezolucije različitim metodama, bilo je potrebno usporediti njihove rezultate. Za usporedbu rezultata, isti model za prepoznavanje teksta primijenjen je na originalnu sliku, sliku visoke rezolucije te sve slike dobivene različitim metodama.

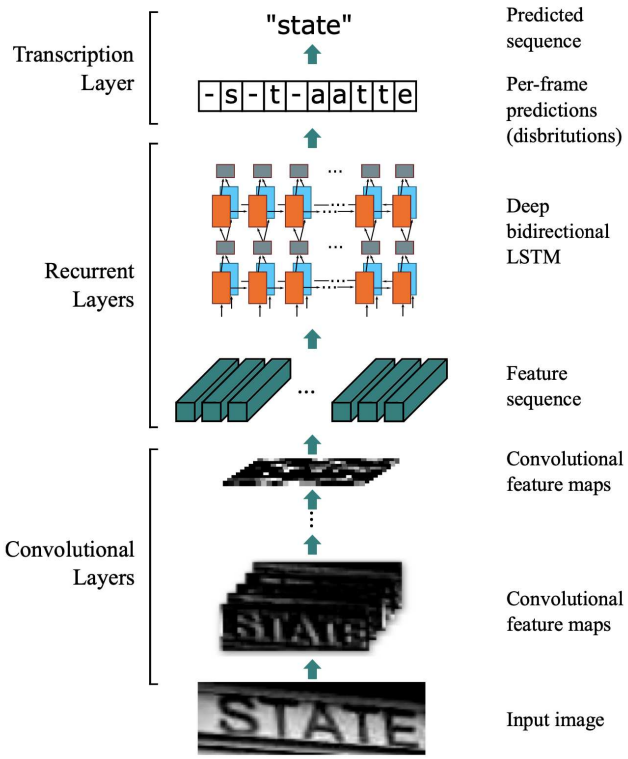
CRNN je često korišten model za prepoznavanje teksta u literaturi [10, 11] zbog svoje robusnosti i sposobnosti za prepoznavanje različitih vrsta teksta. Osim CRNN-a, postoje i drugi modeli za prepoznavanje teksta poput MORAN-a [15] i ASTER-a [16, 17]. CRNN je izabran jer su se metode u literaturi uspoređivale njime te su dostupne najbolje težine dobivene takvom usporedbom.

Konvolucijska rekurentna neuronska mreža (CRNN) je vrsta duboke neuronske mreže posebno dizajnirana za prepoznavanje sekvenci slika, kao što je prepoznavanje teksta u slikama. Ova mreža kombinira konvolucijske slojeve (CNN) za izvlačenje značajki iz slika i rekurentne slojeve (RNN) za modeliranje sekvencijalnih podataka. Arhitektura CRNN-a sastoji se od tri glavna dijela (slika 3.6.):

- Konvolucijski slojevi: Ovi slojevi izvlače vizualne značajke iz ulazne slike. S pomoću filtera koji prolaze po slici te identificiraju obrasce poput tekstura, rubova ili kutova

- Rekurentni slojevi: Prepoznate značajke iz konvolucijskih slojeva se prosljeđuju rekurentnim slojevima. Oni su specijalizirani za obradu sekvencijalnih podataka, poput niza znakova u tekstu. Uzimaju u obzir ne samo trenutnu značajku, već i kontekst prethodnih značajki, što omogućuje razumijevanje redoslijeda znakova i njihovog značenja.

- Transkripcijski sloj: Ovaj sloj je odgovoran za prevođenje izlaza rekurentnih slojeva u konačni tekst. On koristi algoritme poput Connectionist Temporal Classification (CTC) za dekodiranje izlaza mreže i dobivanje prepoznatog teksta.




**Slika 3.6.** Arhitektura CRNN-a [8]

CRNN je posebno učinkovit za prepoznavanje teksta u slikama zbog podržavanja obrade slika različitih veličina te automatskog prilagođavanja duljini teksta. Ključna prednost CRNN-a je mogućnost učenja iz označenih slika. Za učenje nije potrebno raditi segmentaciju znakova ili dodatno procesirati podatke već je moguće učenje direktno korištenjem slika i oznaka teksta.

U radu, CRNN je korišten za usporedbu rezultata različitih metoda. Za svaku metodu, te originalne slike niske i visoke rezolucije, provedena je detekcija teksta s pomoću CRNN modela. Točnost prepoznavanja teksta procijenjena je s pomoću metrike Character Error Rate (CER) [18], čiji je izračun opisan formulom 3.1

$$CER = \frac{(S + D + I)}{(S + D + C)} \tag{3.1}$$

$I$ = minimalan broj znakova koje je potrebno ubaciti

$D$ = minimalan broj znakova koje je potrebno izbrisati

$S$ = minimalan broj znakova koje je potrebno zamjeniti

$C$ = broj točno predviđenih znakova

## 3.4. Bikubna interpolacija [9]

Bikubna interpolacija je jednostavan algoritam za povećanje rezolucije slika. Ona se često koristi za promjenu veličine slika te je izabrana zbog svoje jednostavnosti i dobrih rezultata u odnosu na slične metode (slika 3.7.).


Figure 7. Diagram of nonlinear interpolation algorithm

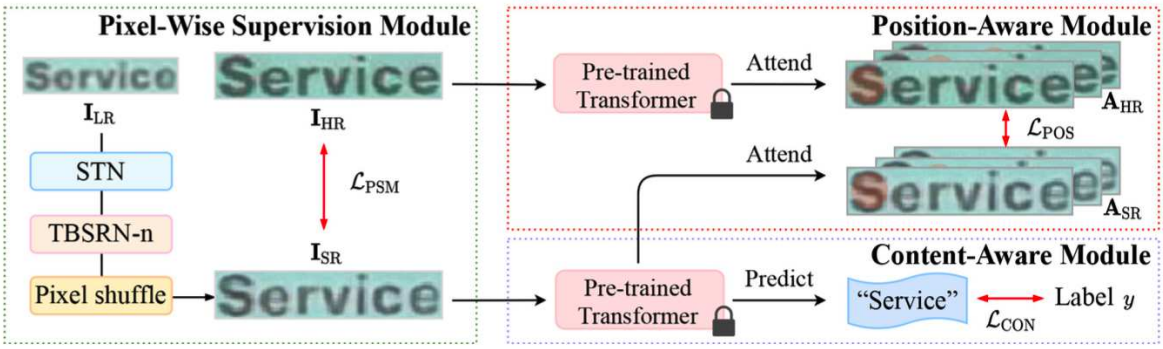
**Slika 3.7.** Bikubna interpolacija u usporedbi s drugim jednostavnim algoritamskim metodama[19]

Za bikubnu interpolaciju korištena je gotova implementacija iz opnecv-python[20] biblioteke. U nastavku je korišteni kod:

```
def upscale_using_bicubic_inteporlation(img):
img = cv2.cvtColor(img, cv2.COLOR_BGR2RGB)
img = cv2.resize(img, (img.shape[1]*2, img.shape[0]*2),
  interpolation=cv2.INTER_CUBIC)
img = cv2.cvtColor(img, cv2.COLOR_RGB2BGR)

return img
```

## 3.5. Scene Text Telescope [10]




**Slika 3.8.** Arhitektura STT-a [10]

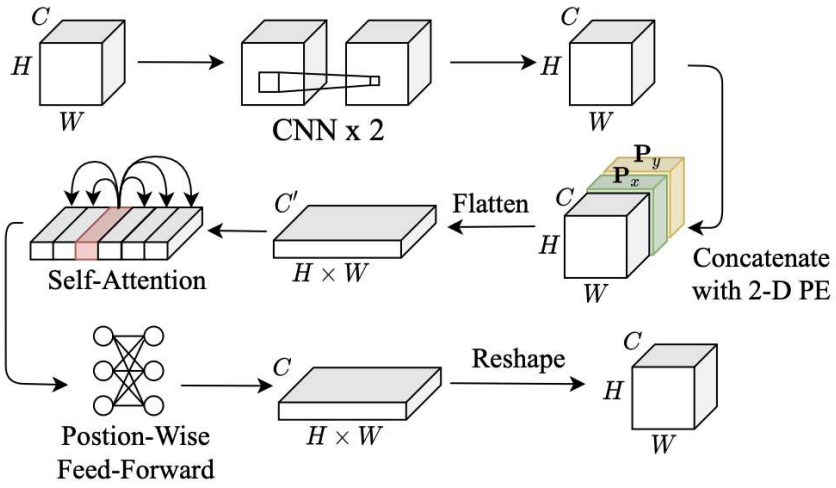
Prvi analizirani model, Scene Text Telescope (STT u nastavku), specifično je dizajniran za poboljšanje rezolucije teksta. Arhitektura STT-a sastoji se od tri glavna modula 3.8. koji svi sudjeluju u procesu učenja, iako samo PWSM modul zapravo generira slike više rezolucije. U nastavku su opisani moduli:

### 3.5.1. Pixel-Wise Supervision Module (PWSM)

PWSM igra ključnu ulogu unutar STT-a. Ovaj modul je odgovoran za povećanje rezolucije te je izoliran s ciljem provođenja eksperimenta. Osmišljen je tako da pri povećanju rezolucije uzima u obzir specifičnosti teksta.

Prvi korak je primjena Spatial Transformer Networka (STN). STN ispravlja geometrijske transformacije u ulaznom tekstu poput nagiba, zakrivljenosti te perspektive kako

bi se slika pripremila za daljnju obradu. Drugi korak je prolazak kroz skup Transformer-Based Super-Resolution Network (TBSRN) blokova (vidi sliku 3.9.). Ovi blokovi koriste kombinaciju CNN-ova, mehanizma pažnje i feed-forward mreže kako bi izvukli i obradili relevantne značajke iz slike te stvorili kartu značajki. U trećem koraku, Pixel shuffle tehnika [21] se primjenjuje na kartu značajki dobivenu u prethodnom koraku kako bi se stvorila slika više rezolucije. Tijekom učenja modela, L2 gubitak između generirane slike visoke rezolucije (SR) i HR slike koristi se za optimizaciju parametara modela.




**Slika 3.9.** Arhitektura TBSRN- a[10]

### 3.5.2. Position Aware Module (PAM)

PAM poboljšava učenje tako što se fokusira samo na točnost teksta, ali ne i rekonstrukciju pozadine slike. Kako bi se to postiglo, korišten je model za prepoznavanje teksta baziran na transformeru, prethodno naučen na umjetnom skupu podataka. Ovaj model generira karte pažnje koje označuju dijelove slike koji su najvažniji za prepoznavanje teksta. Za optimizaciju parametara modela koristi se L1 gubitak između karata pažnje HR i SR slika, čime se potiče model da se fokusira na regije slike koje relevantne za tekst.

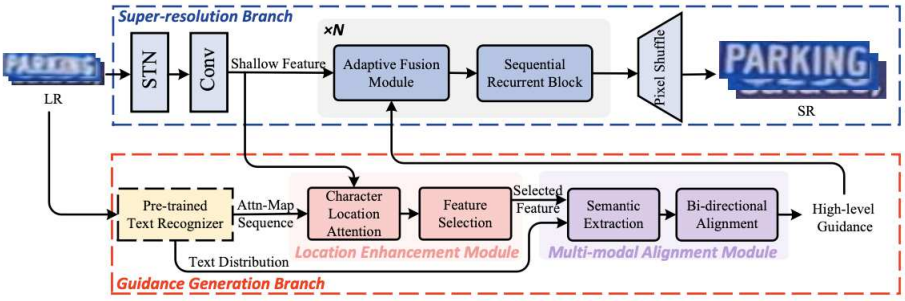
### 3.5.3. Content-Aware Module (CAM)

CAM omogućuje modelu bolje prepoznavanje znakova koji izgledaju slično na slikama poput "c" i "e". Modul, istim modelom za prepoznavanje kao u PAM-u prepoznaje znakove na SR i HR slikama. Zatim koristi Varijacijski Autoenkoder (VAE) za dobivanje latentnog prostora gdje su slični znakovi zajedno grupirani. Pomoću latentnog prostora

računa ponderirani gubitak unakrsne entropije gdje znakovi koji su bliže imaju veću težinu. Time zapravo daje veći značaj greškama sličnih znakova i time navodi model da uči te znakove bolje razlikovati.

## 3.6. Location Enhanced Multi-ModAl Network [11]

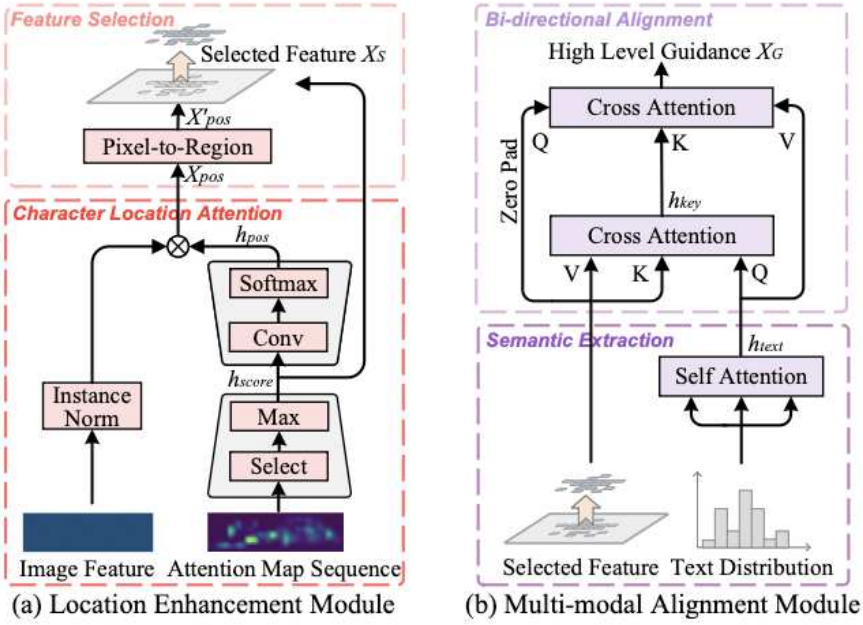
Drugi analizirani model, Location Enhanced Multi-ModAl Network (LEMMA u nastavku) također je specijaliziran za poboljšanje rezolucije slika teksta. Arhitektura modela se može vidjeti na slici 3.10. Za učenje, LEMMA koristi ponderiranu sumu sličnu kao u STT-u. LEMMA model se sastoji od dvije grane opisane u nastavku.




**Slika 3.10.** Arhitektura LEMMA modela [11]

### 3.6.1. Guidance Generation Branch (GGB)

Arhitektura GGB-a, označena crvenim isprekidanim pravokutnikom na slici 3.10., sastoji se od dva glavna modula: Location Enhancement Module (LEM) te Multi-modal Alignment Module (MAM). Njihova arhitektura je vidljiva na slici 3.11.




**Slika 3.11.** Arhitektura LEMMA modula [11]

Slično kao u STT-u, LEM koristi prethodno naučeni model za prepoznavanje teksta za generiranje karata pažnje koje ističu područje s tekstom. S pomoću karata pažnje i dodatnog procesiranja (vidi sliku 3.11.), identificiraju se područja sa znakovima. Zatim se izvlače značajke tih znakova te prosljeđuju dalje MAM-u.

MAM koristi vizualne značajke i semantičke informacije o tekstu koje je prethodno prepoznao model za prepoznavanje teksta. MAM koristi vizualne značajke te semantičke informacije prethodno prepoznate na slici modelom za prepoznavanje teksta. S pomoću tih informacija MAM generira smjernice koje pomažu grani za super-rezoluciju bolje razumijevanje sadržaja teksta.

### 3.6.2. Super-resolution Branch (SRB)

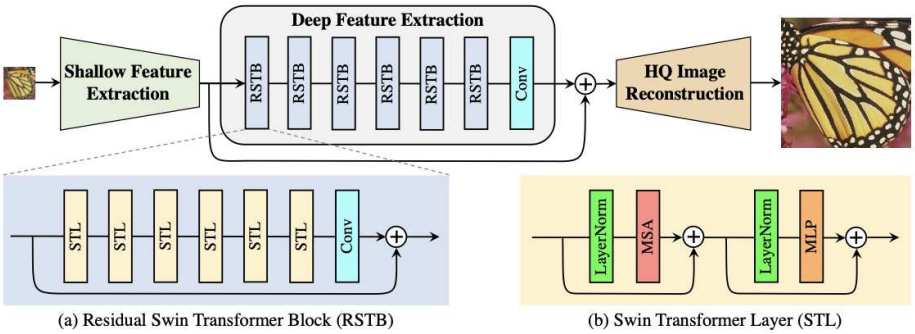
SRB koristi plitke značajke te smjernice generirane u prvoj grani za generiranje slika veće rezolucije. To se postiže nizom koraka koji se sastoje od Adaptive Fusion modula (AFM) te Sequential Recurrent blokova (SRB).

AFM je odgovoran za spajanje plitkih značajki te smjernica iz GGB-a, pripremajući ih za daljnju obradu unutar SRB. AFM prilagođava način spajanja značajka i smjernica osiguravajući samo relevantne informacije u svakom koraku. SRB su zaslužni za pos-

tupno povećanje rezolucije slike kroz više koraka. Konačno, Pixel shuffle tehnikom se kreiraju nove slike.

## 3.7. SwinIR [12]

Treći analizirani model, SwinIR, temelji se na Swin Transformerima. SwinIR model je namijenjen primarno za restauraciju slika što uključuje povećanje rezolucije, uklanjanje šuma te smanjenje kompresijskih artefakata. Iako SwinIR nije specifično dizajniran za slike teksta, u ovom radu analiziramo njegovu učinkovitost kao općenitijeg modela u usporedbi sa specijaliziranim modelima za tekst. Model se sastoji od 3 glavna modula (prikazana na slici 3.12.), koji su opisani u nastavku.




**Slika 3.12.** Arhitektura SwinIR-a [12]

### 3.7.1. Shallow Feature Extraction (SFE)

SFE modul koristi konvolucijski sloj kako bi izvukao plitke značajke slike poput rubova. Izvučene značajke se prosljeđuju sljedećim modulima na daljnju obradu.

### 3.7.2. Deep Feature Extraction (DFE)

DFE modul izvlači dublje, apstraktnije značajke sa slika. Modul se sastoji od niza Residual Swin Transformer blokova (RSTB), prikazanih na slici 3.12. Na kraju modula nalazi se konvolucijski sloj. Svaki RSTB se temelji na Swin Transformer Layerima (slika 3.12. pod (a)) koji koriste mehanizam pažnje te pomični prozor kako bi modelirali lokalne i globalne ovisnosti u slici. Svaki RSTB također završava konvolucijskim slojem.

### 3.7.3. High Quality Image Reconstruction (HQIM

Zadnji modul je HQIR koji koristeći značajke izvučene u prethodna dva modula se spajaju kako bi se rekonstruirala slika. Postupak rekonstrukcije ovisi o specifičnom zadatku. Za zadatke povećanje rezolucije koristi se transponirana konvolucija nad dobivenim značajkama čime se generiraju novi detalji u slici i povećava rezolucija slike. Za zadatke koji ne zahtijevaju povećanje rezolucije, poput uklanjanja šuma ili kompresijskih artefakata, koristi se samo konvolucijski sloj za obradu značajki.

## 3.8. Dodatni modeli za povećanje rezolucije teksta

Osim navedenih, za rad su razmatrane dodatne implementacije modela, ali zbog raznih ograničenja nisu korišteni u radu. Prva razmatrani model je bio TATT [22]. TATT nije korišten zbog nedostupnosti prethodno naučenih težina modela i očekivanog dugog vremena učenja. CCD [23] je također razmatran. Unatoč dostupnosti težina za CCD, nije uspješno pokrenut te je zato izostavljen. Text Gestalt [24] je bio također razmatran, ali prethodno naučene težine modela nisu bile dostupne bez posebnog računa kojeg nije bilo moguće kreirati iz Hrvatske.

# 4. Organizacija i pokretanje eksperimenata

Većina implementacija modela dohvaćena je s GitHub repozitorija autora originalnih radova. U većinu implementacija dodan je kod koji omogućuje spremanje rezultata povećanja rezolucije. Time je omogućena usporedba kvalitete povećanja rezolucije različitih modela neovisno o modelu za prepoznavanje teksta korištenih u implementacijama.

## 4.1. Organizacija repozitorija

Organizacija repozitorija za lakšu reprodukciju rezultata.

- Dataset - Sadrži skup podataka i skripte za obradu podataka..
  - TextZoom - Originalni skup podataka u obliku mdb datoteka.
  - TextZoom_unpacked - Skup podataka samo slike teksta bez specijalnih znakova te rezultate eksperimenata.
  - own - Skup podataka sastavljen od isječaka vlastitih slika.
  - own_images - Vlastite slike korištene za vlastiti skup.
  - unpacking_images.ipynb - Jupyter bilježnica za izvlačenje LR i HR slika iz mdb formata te spremanje u njega.
  - accuracy.py - Implementacija CER funkcije.
  - calculate_accuracy-ipynb - Jupyter bilježnica za računanje točnosti eksperimenata.

  - create_bicubic - Jupyter bilježnica za povećanje rezolucije slika bikubnom interpolacijom.
  - visualise_results - Jupyter bilježnica za vizualizaciju rezultata.
- Implementations - Implementacije eksperimenata
  - FundanOCR - Repozitorij originalno s više implementacija super rezolucije, sadrži implementaciju STT.
  - crnn.pytorch - Repozitorij s implementacijom CRNN-a.
  - LEMMA - Repozitorij s LEMMA implementacijom.
  - numpy_vs_vanilla.py - Pokazni isječak koda za brzinu numpy-a.
  - SwinIR.ipynb - Bilježnica za pokretanje SwinIR modela.

## 4.2. Pokretanje eksperimenata

U ovom poglavlju opisane su naredbe korištene za pokretanje eksperimenata s ciljem mogućnosti reprodukcije rezultata iz rada. Prije svakog od eksperimenata potrebno je instalirati sve potrebne biblioteke pomoću "requirements.txt" datoteka.

### 4.2.1. STT

STT eksperiment je pokrenut idućom naredbom

```
1      python main.py --batch_size=16 --STN --exp_name SR --text_focus
   --super_resolution --super_resolution_dir ./super_resolution
   --resume./checkpoint/model_best.pth
```

U ovoj naredbi SR je ime eksperimenta. Prije pokretanja, potrebno je stvoriti direktorij s istim nazivom unutar checkpoint direktorija te unutar njega log.txt datoteku. Originalni repozitorij nije imao opciju za spremanje generiranih slika, zato je dodana opcija –super_resolution LR slike čitaju se iz direktorija navedenog pod –super_resolution_dir. Unutar istog direktorija stvara se novi imena STT u koji se spremaju generiranje slike.

### 4.2.2. LEMMA

Unutar LEMMA eksperimenta, konfiguracija se provodi kroz "super_resolution.yaml" datoteku. Unutar nje postavljaju se putanje do skupa podataka te težine modela. U ovom eksperimentu, slike se čitaju iz mdb datoteka. Pokretanjem, unutar repozitorija stvaraju se direktoriji "easy", "medium" i "hard". Unutar njih su slike povećane rezolucije. Nakon provođenja eksperimenta potrebno je dodatno ukloniti slike s tekstom koji sadrži specijalne znakove. Eksperiment je pokrenut idućom naredbom, gdje –test označava test način rada.

```
1    python main.py --test
```

### 4.2.3. SwinIR

Za pokretanje SwinIR implementacije korištena je Google Colab bilježnica autora rada. Bilježnica je promijenjena kako bi omogućila izvođenje samo SwinIR eksperimenta iako originalno sadrži dodatne implementacije. Za pokretanje je potrebno pokrenuti sve ćelije redom. Bilježnica je pokretana unutar Google Colab okruženja.

# 5. Rezultati

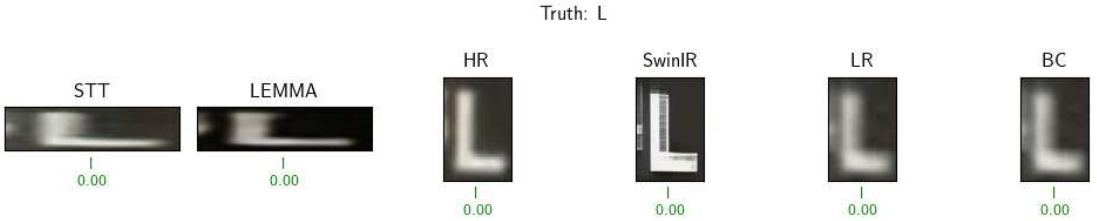
Uspoređivana je jedna algoritamska metoda (bikubna interpolacija), jedna metoda temeljena na dubokom učenju koja nije specifično dizajnirana za povećanje rezolucije teksta (SwinIR) te dvije metode specijalizirane za povećanje rezolucije teksta (LEMMA te STT). SwinIR je uključen u usporedbu kako bi se procijenila razlika između specijaliziranih i općenitijih metoda.

Zbog ograničenih računalnih resursa, korištene su prethodno naučene težine objavljene od strane autora. Budući da su LEMMA i STT već učeni na TextZoom skupu podataka, dodatno učenje nad istim skupom vjerojatno ne bi značajno promijenilo rezultate nad tim skupom. Međutim, dodatno učenje bi vjerojatno utjecalo na rezultate nad vlastitim skupom podataka. Rezultate različitih metoda uspoređujemo na tri podskupa TextZoom skupa podataka (easy, medium i hard) te vlastitom skupu podataka.

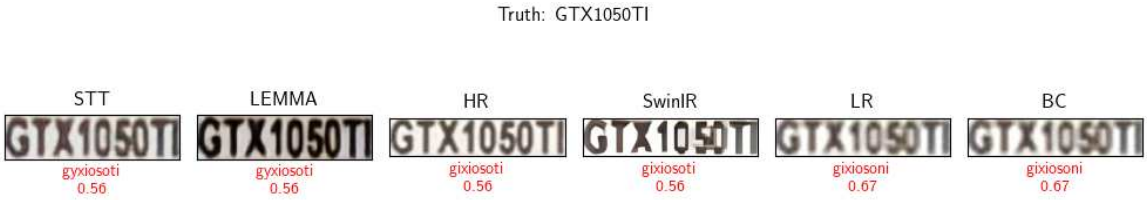
Primjeri rezultata iz različitih podskupova su na slikama u nastavku. Za svaku sliku prikazana je originalna slika niske rezolucije (LR), originalna slika visoke rezolucije (HR) te četiri verzije slike dobivene povećanjem rezolucije različitim metodama. Iznad svakog skupa slika prikazan je referentni tekst, dok ispod svake od slika se nalazi tekst prepoznat s pomoću CRNN modela s pripadajućom CER vrijednošću.

## 5.1. Easy

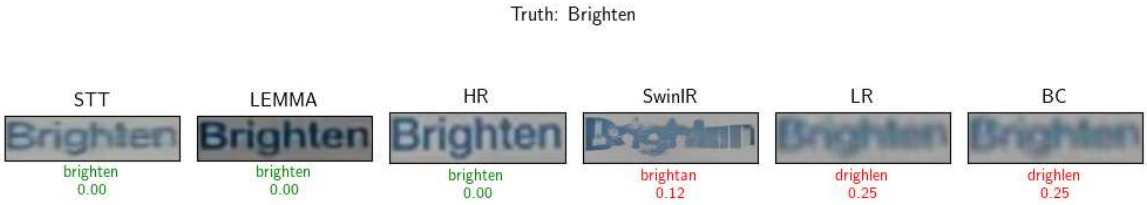
Primjeri rezultat iz podskupa prikazani su na slikama 5.1., 5.2., 5.3., 5.4., 5.5., 5.6. te 5.7.



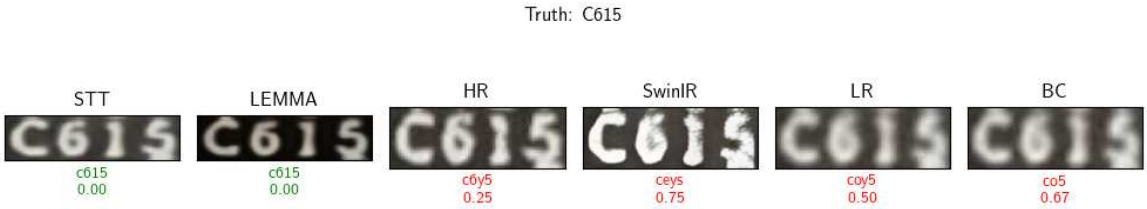

**Slika 5.1.** Primjer gdje su sve metode uspješne.




**Slika 5.2.** Primjer gdje su sve metode neuspješne

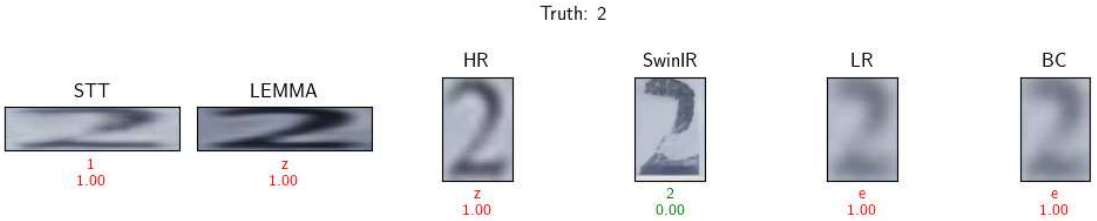



**Slika 5.3.** Primjer gdje su metode specijalizirane za tekst uspješne

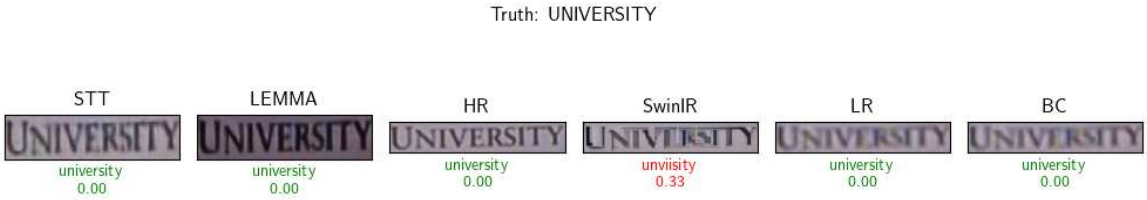



**Slika 5.4.** Primjer gdje su metode specijalizirane za tekst uspješnije od slike više rezolucije


**Slika 5.5.** Primjer gdje je metoda koja nije specijalizirana za tekst uspješna s jednom od metoda specijaliziranih za tekst




**Slika 5.6.** Primjer gdje je metoda koja nije specijalizirana za tekst najuspješnija




**Slika 5.7.** Primjer gdje je metoda koja nije specijalizirana za tekst pogoršala rezultat

Easy podskup podataka sadrži 1462 slike. Prikazani primjeri ilustriraju raznovrsnost rezultata i nisu nužno reprezentativni za ukupnu distribuciju rezultata u skupu. U nekim slučajevima, povećanje rezolucije nije bilo ni potrebno da bi tekst bio uspješno prepoznat (npr. slika 5.1.), dok u drugim slučajevima neovisno o metodi tekst nije bio uspješno prepoznat (npr. slika 5.2.). U većini slučajeva (npr. slike 5.3. i 5.4.) metode dubokog učenja značajno poboljšavaju čitljivost teksta. Posebice je zanimljiv SwinIR model. Unatoč tome što nije specijaliziran za tekst, SwinIR ponekad postiže bolje rezultate od svih ostalih metoda (npr. slika 5.6.). Takvi primjeri su rijetki i češći su primjeri gdje pokazuje lošije rezultate od svih ostalih metoda, uključujući i od LR (npr. slika 5.7.). Uočljivo je da, iako SwinIR ponekad poboljšava čitljivost teksta, te slike uglavnom ne izgledaju poput prirodnog teksta već sadrže artefakte i teksture koje ne nalikuju na tekst.

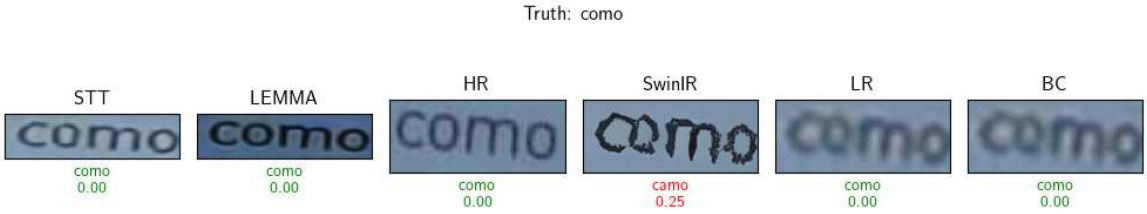
Prosječne CER vrijednosti za sve metode te LR i HR slike prikazane su u tablici 5.1. Rezultati pokazuju da HR slike imaju najnižu grešku što je i očekivano. LEMMA i STT modeli značajno poboljšavaju točnost prepoznavanja teksta u odnosu na LR slike, dok razlika između LR i BC slika je minimalna. SwinIR, iako ponekad uspješan, u prosjeku ima najveću grešku, veću i od LR slika.

Tablica 5.1. CER vrijednosti experimenata za easy skup

| LR | HR | BC | SwinIR | STT | LEMMA |
|---|---|---|---|---|---|
| 0.3979 | **0.1096** | 0.3864 | 0.4452 | 0.2259 | 0.1697 |

## 5.2. Medium

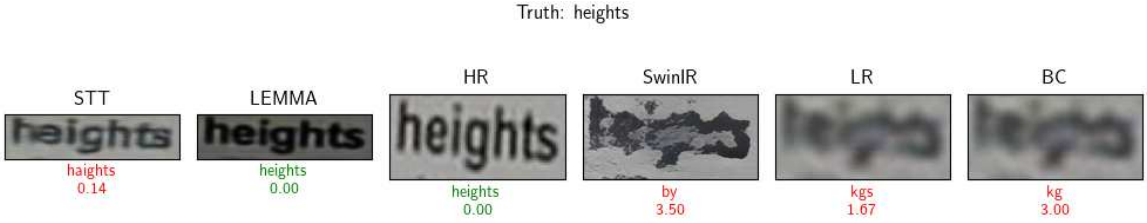
Primjeri iz medium podskupa prikazani su na slikama: 5.8., 5.9., 5.10. te 5.11.




**Slika 5.8.** Primjer gdje su sve metode osim SwinIR uspješne


**Slika 5.9.** Primjer gdje su metode specijalizirane za tekst uspješne


**Slika 5.10.** Primjer gdje je STT najuspješniji




**Slika 5.11.** Primjer gdje su LEMMA i HR uspješni

Medium podskup sadrži 1241 sliku. U ovom podskupu podataka, modeli specijalizirani za tekst pokazuju svoju prednost. Uglavnom daju značajno bolje rezultate od drugih metoda te u nekim slučajevima (npr. slika 5.10.) pokazuju i bolje rezultate od HR slika. Unutar ovog podskupa, ograničenja SwinIR modela postaju očita. Većina generiranih rezultata nalikuju na mrlje (npr. slike 5.11. i 5.9.), umjesto na tekst.

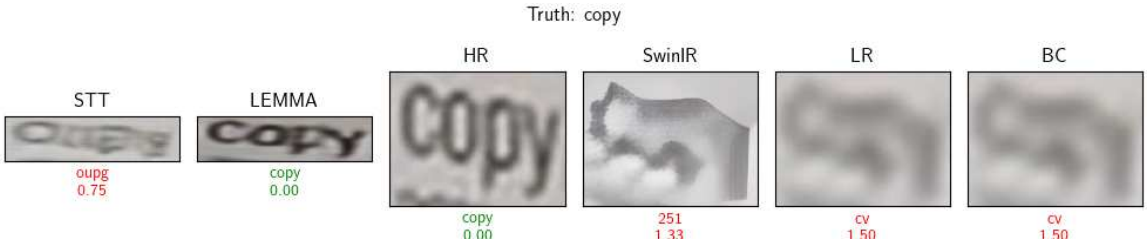
Tablica 5.2. pokazuje kako je ovaj podskup sadržavao teže slike koje su dovodile do lošijih rezultata. Unutar ovog podskupa, razlika između STT i LEMMA modela se znatno povećala te potvrđuje veću uspješnost LEMMA modela. Razlika između SwinIR te LR se smanjila, ali unatoč tome SwinIR još uvijek postiže lošije rezultate. Razlika između BC i LR slika je opet minimalna.

Tablica 5.2. CER vrijednosti eksperimenata za medium skup

| **LR** | **HR** | **BC** | **SwinIR** | **STT** | **LEMMA** |
|---|---|---|---|---|---|
| 0.5611 | **0.1287** | 0.5598 | 0.5856 | 0.3496 | 0.2407 |

## 5.3. Hard

Primjeri iz hard podskupa prikazani su na slikama: 5.12., 5.13., 5.14. te 5.15.




**Slika 5.12.** Primjer gdje su LEMMA i HR čitljivi


**Slika 5.13.** Primjer gdje je samo HR čitljiv


**Slika 5.14.** Primjer loše slike gdje je tekst nečitljiv u svim slučajevima


**Slika 5.15.** Primjer gdje je STT jedini gdje tekst nije mogao biti pročitan

Hard podskup podataka sadrži 1207 slika. Hard podskup se pokazao najzahtjevniji za CRNN model u usporedbi s ostalim podskupovima iz TextZoom skupa podataka. Tablica 5.3. pokazuje znatno lošije rezultate prepoznavanja za sve metode. Značajno pogoršanje rezultata uočeno je i kod STT i LEMMA modela, što pokazuje osjetljivost na lošiju kvalitetu slika prisutnu u podskupu. Na primjer, na slici 5.14. tekst na LR slici je jedva čitljiv, što rezultira lošim rezultatom svih metoda. Na istoj slici (slika 5.14.) se vidi da su STT i LEMMA uspjeli rekreirati nekakav tekst, ali zbog loše kvalitete LR slike tekst koji je rekreiran se ne poklapa s pravim. SwinIR i dalje postiže iznimno loše rezultate, lošije nego LR slike. Bikubna interpolacija i dalje daje rezultate slične LR slikama, iako ovaj put s neznatno većom greškom.

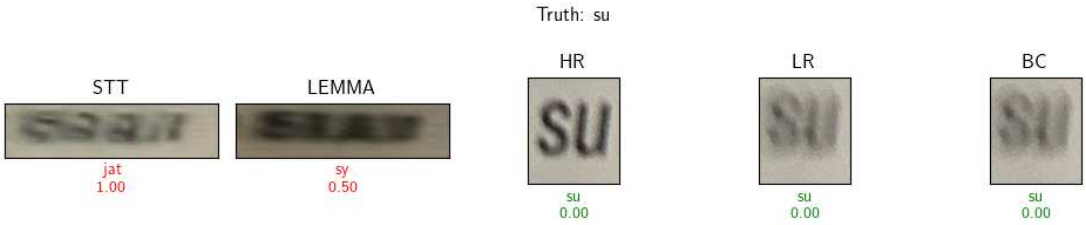
Tablica 5.3. CER vrijednosti eksperimenata za hard skup

| **LR** | **HR** | **BC** | **SwinIR** | **STT** | **LEMMA** |
|---|---|---|---|---|---|
| 0.6056 | **0.2059** | 0.6074 | 0.6351 | 0.4742 | 0.3864 |

## 5.4. Vlastite slike

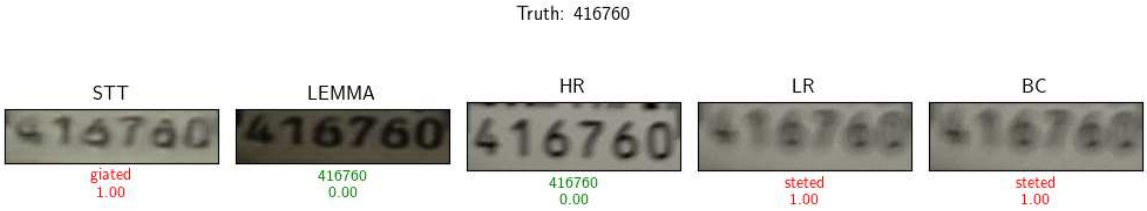
Eksperimenti su također proveden na vlastitom skupu podataka, izuzev SwinIR koji je zbog nezadovoljavajućih rezultata na TextZoom skupu izostavljen. Primjeri iz vlastitog

skupa prikazani su na slikama: 5.16., 5.17., 5.18. i 5.19.




**Slika 5.16.** Primjer gdje su LEMMA i STT pogoršali predikciju


**Slika 5.17.** Primjer gdje je sve teško čitljivo




**Slika 5.18.** Primjer gdje su LEMMA i HR dobri


**Slika 5.19.** Primjer gdje su STT i LEMMA uspješni

Rezultati na vlastitom skupu razlikuju se od rezultata na TextZoom skupu. Primjerice, na slici 5.16., obje specijalizirane metode povećanja rezolucije (LEMMA i STT) pogoršale su sposobnost čitanja teksta u odnosu na LR, što je bila rijetkost na TextZoom skupu. Naravno, postoje i slučajevi gdje su te metode uspješne (npr. slika 5.19.). Zanimljivo je da se trendovi uočeni na TextZoom skupu ne ponavljaju na vlastitom skupu.

Tablica 5.4. pokazuje da su obje specijalizirane metode, u prosjeku, dovele do povećanja CER vrijednosti što znači da su pogoršale čitljivost teksta umjesto olakšale.

Tablica 5.4. CER vrijednosti eksperimenata za vlastiti skup

| **LR** | **HR** | **BC** | **STT** | **LEMMA** |
|---|---|---|---|---|
| 0.6835 | **0.1119** | 0.6835 | 0.8560 | 0.7815 |

Mogućih objašnjenja ovakvih rezultata ima nekoliko. Prvo, vlastiti skup se sastoji od svega 12 slike što označava statističku značajnost rezultata. Drugo, skup sadrži slike s velikom razlikom između LR i HR slika (npr. slika 5.17.). U takvim slučajevima, činjenica da CRNN model da prepoznaje tekst na HR slici, ali ne i na ostalim slikama ne ukazuje nužno na loše performanse metoda, već na teške uvjete prepoznavanja na tim slikama. Ograničen broj slika u skupu je dijelom i poslije nekih iznimno loših LR slika koje su morale biti uklonjene jer nisu bile prepoznatljive ni ljudskom oku. Dodatni problem bi mogle predstavljati različite veličine slika. Unutar TextZoom skupa sve slike su bile do svega 50x50 piksela dok u vlastitom skupu postoje slike vrlo različitih dimenzije, gdje neke dosežu veličine od nekoliko stotina piksela (poput 5.19.). Za dublju analizu i identifikaciju uzroka ovih rezultata bilo bi potrebno kreirati veći i raznovrsniji skup podataka, što nažalost nije bilo moguće u okviru ovog rada.

## 5.5. Ukupno

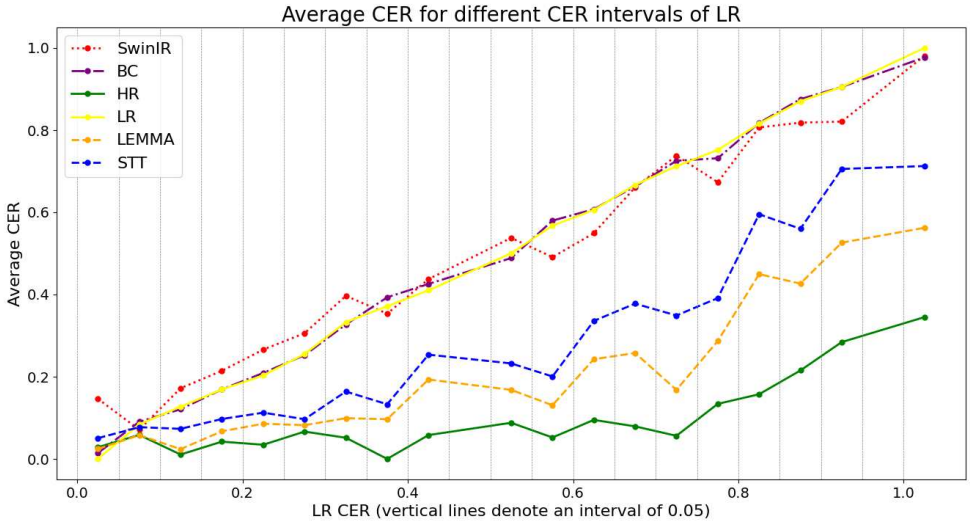
U tablici 5.5. se vide rezultati svih eksperimenata na svim skupovima.

Tablica 5.5. CER vrijednosti eksperimenata za sve skupove

| **Skup podataka** | **LR** | **HR** | **BC** | **SwinIR** | **STT** | **LEMMA** |
|---|---|---|---|---|---|---|
| **easy** | 0.3979 | **0.1096** | 0.3864 | 0.4452 | 0.2259 | 0.1697 |
| **medium** | 0.5611 | **0.1287** | 0.5598 | 0.5856 | 0.3496 | 0.2407 |
| **hard** | 0.6056 | **0.2059** | 0.6074 | 0.6351 | 0.4742 | 0.3864 |
| **vlastiti** | 0.6835 | **0.1119** | 0.6835 | - | 0.8560 | 0.7815 |

Unutar TextZoom skupa, STT i LEMMA su uspješno poboljšali čitljivost teksta. SwinIR je davao lošije rezultate od LR slike te time pokazao da za povećanje rezolucije slika s ciljem lakšeg čitanja teksta nije dovoljno koristiti nasumični model za povećanje rezolucije već je potrebno koristiti model specijaliziran za zadatak. Uspješnost SwinIR bi se vjerojatno poboljšala dodatnim učenjem nad TextZoom skupom, ali to nije evaluirano u sklopu rada.

Slika 5.20. prikazuje odnos između kvaliteta originalnih LR slika te uspješnosti različitih metoda poboljšanja rezolucije analiziranih u ovom radu. Prosječne CER vrijednosti računate su grupiranjem rezultata prema CER vrijednosti LR slika. Na primjer, prva točka na grafu za svaku metodu predstavlja prosječnu CER vrijednost slika dobivenih tim metodama čija je originalna CER vrijednost (CER vrijednost LR slike) bila između 0 i 0.05. Graf prikazuje samo vrijednosti do 1.05, podijeljene na intervale od 0.05, jer je broj slika s većom CER vrijednošću bio premalen za analizu.




**Slika 5.20.** Graf koji predstavlja prosječnu CER vrijdnost metoda ovisno o CER vrijednosti pripadajuće LR slike (TextZoom)

Graf pokazuje da greška slika generiranih LEMMA i STT modelima počinje naglije rasti kada CER vrjednost LR slike prelazi približno 0.6. Do te vrijednost, oba modela značajno poboljšavaju čitljivost s prosječnom CER vrijednošću manjom od 0.2. LEMMA model značajno dulje zadržava vrijednost manju od 0.2 čak i kada CER vrijednost LR slike počinje prelaziti 0.7, što ukazuje na veću otpornost modela na lošiju kvalitetu slika.

Svukupno, na jednostavijim slikama i LEMMA i STT ostvaruju odlične rezultate, ali kako slike postaju lošije kvalitete, LEMMA model povećava svoju prednost nad STT modelom.

Na grafu se uočava da bikubna interpolacija u prosjeku ne olakšava prepoznavanje teksta u odnosu na LR slike. To je očekivano jer CRNN u svojoj implementaciji koristili bilinearnu interpolaciju za promjenu veličina slike, a rezultati bilinearne i bikubne interpolacije su u mnogim slučajevima veoma slični.

# 6. Zaključak

Unutar rada uspješno su uspoređena dva modela (STT i LEMMA) napravljena s ciljem povećanja rezolucije teksta, bikubna interpolacija te model za povećanje rezolucije namijenjen za općenitiju uporabu (SwinIR). Poseban fokus je bio nad STT i LEMMA modelima te njihovoj mogućnosti poboljšavanja čitljivosti teksta. Rezultati eksperimenata na TextZoom skupu podataka pokazali su da specijalizirani modeli, posebno LEMMA, značajno poboljšavaju čitljivost teksta na slikama loše kvalitete u odnosu na ostale metode. Također, pokazano je kako općenitija metoda povećanja rezolucije često radi na štetu čitljivosti teksta. Međutim, učinkovitost metoda znatno pada pogoršavanjem kvalitete ulaznih slika, toliko da na slikama vrlo loše rezolucije često i otežavaju čitanje originalnog teksta. Takvi rezultati ukazuju na potrebnu za daljnjim istraživanjem s ciljem razvitka robusnijih metoda koje se mogu nositi s ulaznim slikama vrlo loše kvalitete.

Eksperimenti na vlastitom skupu podataka, iako ograničeni veličinom i raznolikošću, ukazali su na potencijalne probleme u generalizaciji modela na slike koje se razlikuju od onih na kojima su trenirani. Također potencijalno ukazuju na limitiranu uspješnost modela na slikama iznimno niske rezolucije. Ti rezultati naglašavaju važnost testiranja modela na raznolikim skupovima podataka kako bi se osigurala njihova praktična primjena.

Unutar rada obuhvaćene su samo dvije metode specijalizirane za tekst, unatoč velikom broju literature na temu. Iz tog razloga, rezultati ovog rada nisu nužno reprezentacija najboljih rezultata postignutih u području, već pokazuju mogućnosti modernih metoda te smjer u kojem se istraživanje razvija. Poboljšanje rezolucije ostaje vrlo aktivno područje koje će postajati samo još relevantnije s vremenom te se očekuje razvoj još boljih metoda u skorije vrijeme.

# Literatura


[1] Python, https://www.python.org/about/, [mrežno; stranica posjećena: svibanj 2024.].

[2] PyTorch Foundation, https://pytorch.org/docs/stable/index.html, [mrežno; stranica posjećena: svibanj 2024.].

[3] NumPy, https://numpy.org, [mrežno; stranica posjećena: svibanj 2024.].

[4] NVIDIA, https://developer.nvidia.com/cuda-zone, [mrežno; stranica posjećena: svibanj 2024.].

[5] Jupyter, https://jupyter.org, [mrežno; stranica posjećena: svibanj 2024.].

[6] Google, https://research.google.com/colaboratory/faq.html, [mrežno; stranica posjećena: svibanj 2024.].

[7] W. Wang, E. Xie, X. Liu, W. Wang, D. Liang, C. Shen, i X. Bai, "Scene text image super-resolution in the wild", u *Computer Vision – ECCV 2020*, A. Vedaldi, H. Bischof, T. Brox, i J.-M. Frahm, Ur. Cham: Springer International Publishing, 2020., str. 650–666. https://doi.org/10.1007/978-3-030-58607-2_38

[8] B. Shi, X. Bai, i C. Yao, "An end-to-end trainable neural network for image-based sequence recognition and its application to scene text recognition", *IEEE Trans. Pattern Anal. Mach. Intell.*, sv. 39, br. 11, str. 2298–2304, 2017.

[9] geeksforgeeks, https://www.geeksforgeeks.org/python-opencv-bicubic-interpolation-for-resizing-image/, [mrežno; stranica posjećena: svibanj 2024.].

[10] J. Chen, B. Li, i X. Xue, "Scene text telescope: Text-focused scene image super-resolution", u *CVPR*, 2021., str. 12 026–12 035.

[11] H. Guo, T. Dai, G. Meng, i S.-T. Xia, "Towards robust scene text image super-resolution via explicit location enhancement", u *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence, IJCAI-23*. International Joint Conferences on Artificial Intelligence Organization, 8 2023., str. 782–790, main Track. https://doi.org/10.24963/ijcai.2023/87

[12] J. Liang, J. Cao, G. Sun, K. Zhang, L. Van Gool, i R. Timofte, "Swinir: Image restoration using swin transformer", *arXiv preprint arXiv:2108.10257*, 2021.

[13] Uni Matrix Zero, https://unimatrixz.com/topics/ai-upscaler/upscaling-methods/, [mrežno; stranica posjećena: svibanj 2024.].

[14] Wenjia Wang, https://github.com/WenjiaWang0312/TextZoom/blob/master/easy_medium_hard.jpg, [mrežno; stranica posjećena: svibanj 2024.].

[15] C. Luo, L. Jin, i Z. Sun, "Moran: A multi-object rectified attention network for scene text recognition", *Pattern Recognition*, sv. 90, str. 109–118, 2019.

[16] B. Shi, M. Yang, X. Wang, P. Lyu, C. Yao, i X. Bai, "Aster: An attentional scene text recognizer with flexible rectification", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, sv. 41, br. 9, str. 2035–2048, 2019.

[17] B. Shi, X. Wang, P. Lyu, C. Yao, i X. Bai, "Robust scene text recognition with automatic rectification", u *2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, June 27-30, 2016*, 2016., str. 4168–4176.

[18] OECD, https://oecd.ai/en/catalogue/metrics/character-error-rate-%28cer%29, [mrežno; stranica posjećena: svibanj 2024.].

[19] D. Han, "Comparison of commonly used image interpolation methods", u *2nd International Conference on Computer Science and Electronics Engineering (ICCSEE)*, sv. 2. Atlantis Press, 2013., str. 1556–1559.

[20] opencv, https://pypi.org/project/opencv-python/, [mrežno; stranica posjećena: svibanj 2024.].

[21] W. Shi, J. Caballero, F. Huszár, J. Totz, A. P. Aitken, R. Bishop, D. Rueckert, i Z. Wang, "Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network", 2016.

[22] J. Ma, Z. Liang, i L. Zhang, "A text attention network for spatial deformation robust scene text image super-resolution", u *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022., str. 5911–5920.

[23] T. Guan, W. Shen, X. Yang, Q. Feng, Z. Jiang, i X. Yang, "Self-supervised character-to-character distillation for text recognition", u *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, October 2023., str. 19 473–19 484.

[24] J. Chen, H. Yu, J. Ma, B. Li, i X. Xue, "Text gestalt: Stroke-aware scene text image super-resolution", 2021.

# Sažetak


## Poboljšanje rezolucije slika teksta

Jura Hostić

Prepoznavanje teksta je zadatak sa širokim spektrom primjena poput digitalizacije dokumenata ili autonomne vožnje. Mnoge od modernih metoda ne postižu zadovoljavajuće rezultate nad slikama lošije rezolucije što naglašava važnost tehnika povećanja rezolucije. Cilj ovoga rada je usporediti algoritamske metode s općenitim i specijaliziranim metodama dubokog učenja te procijeniti njihovu uspješnost. Uspoređene su dvije specijalizirane metode (LEMMA i STT), jedna općenita metoda (SwinIR) te bikubna interpolacija kao algoritamska metoda. Uspješnost svake metode evaluirana je s pomoću CRNN modela za čitanje teksta nad podskupu TextZoom skupa podataka te vlastitom skupu podataka. Rezultati pokazuju kako duboki modeli opće namjene pogoršavaju čitljivost originalnog teksta dok specijalizirane metode značajno poboljšavaju točnost prepoznavanja teksta. Također, rezultati ukazuju da je učinkovitost svih metoda ograničena kvalitetom originalnih slika.



**Ključne riječi:** duboko učenja; povećanje rezolucije slika; prepoznavanje teksta; konvolucijske neuralne mreže

# Abstract


## Improving resolution of text images

Jura Hostić

Text detection is a task with a wide range of applications, such as document digitalization or autonomous driving. Many modern methods do not achieve satisfactory results on low resolution images, which emphasizes the importance of super-resolution techniques. The goal of this paper is to compare algorithmic methods with general and specialized deep learning methods and evaluate their success. Two specialized methods (LEMMA and STT), one general method (SwinIR), and bicubic interpolation as an algorithmic method were compared. The success of each method was evaluated using the CRNN text recognition model on a subset of the TextZoom dataset as well as on a custom dataset. The results show that general deep learning models worsen the readability of the original text, while specialized methods significantly improve the accuracy of text recognition. Also, the results indicate that the effectiveness of all methods is limited by the quality of the original images.



**Keywords:** deep learning; super-resolution; text recognition; convolutional neural networks