

Семинарска работа по предметот Дигитално процесирање на слика

Тема: Боење на црно-бели слики

Изработила: Марија Цветаноска

Ментор: д-р Ивица Димитровски

Вовед

Колоризација на слика, односно боење на црно-бели слики, е процесот на земање како влез црно-бела (grayscale) слика и добивање како излез слика во боја која ги репрезентира семантичките бои и тонови на влезната слика (на пример, океанот на сончево време треба да има сина боја – не смее да биде обоен розев од страна на моделот). Постојат повеќе методи и пристапи кон оваа проблематика, но овде ќе се задржиме и ќе ги обработиме алгоритмите и методите кои се елаборирани во научниот труд на Zhang et al., *Colorful Image Coloration*, објавен во 2016 година на Европската Конференција за Компјутерска Визија (ECCV).

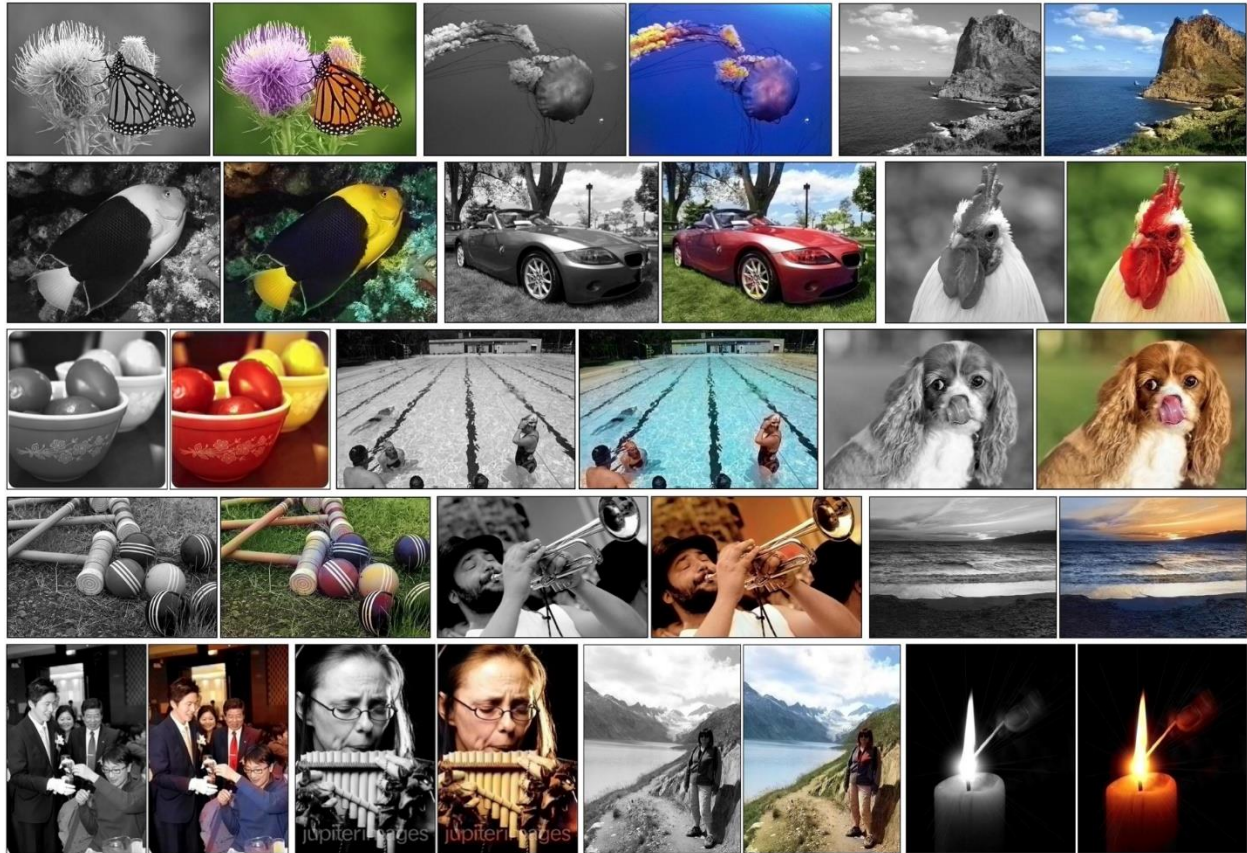
Претходните пристапи кон колоризација на црно-бели слики се засновале на рачни аотирања од страна на луѓето со кои често се добивале бледи бои како резултат и сликите не изгледале „реални“, односно било воочливо дека се добиени како резултат при колоризација на црно-бели слики.

Пристапот на Zhang кон овој проблем се заснова на тоа дека овој проблем треба да се реши со методите на Длабоко учење, односно со користење на Конволуциски невронски мрежи (CNN) со кои ќе се „халуцинира“, т.е. предвиди како треба да изгледа една црно-бела слика кога би била во боја. Овој метод се докажал како успешен и како резултат дава реалистични и живописни колоризации на сликите.

Идејата на Zhang за колоризација на слики

Кога ќе се погледнат црно-белите слики на Слика 1, на прв поглед изгледа претешко да се исхалуцинираат боите на сликата, со оглед на тоа што голем дел од информациите (две од вкупно три димензии) се изгубени. Но кога ќе ги погледнеме повнимателно ќе забележиме дека во повеќето случаи, семантиката на една сцена и текстурите на површините се значително воочливи за повеќето слики: тревата е вообичаено зелена, небото сино, а бубамарата е сигурно црвена. Се разбира, овие семантички знаци не секогаш се точни и не секогаш функционираат добро во секој случај. На пример, топчињата на тревата може да не се црвени, жолти или виолетови (како на резултатот од колоризацијата). Но сепак, ова се смета за добра предикција. Сепак, идејата за колоризацијата не е да се обојат сликите како во нивната реалност, туку да се продуцира пригодна и веродостојна колоризација која би можела да го излаже човечкото око и да се мисли дека сликата е оригинално фотографирана во боја.

Бојење на црно-бели слики



Слика 1 - Неколку примери за фотографии колоризирани со овој алгоритам

Методот на Zhang за колоризација на црно-бели слики започнува со конвертирање на сликата што сакаме да ја обоиме од RGB во Lab просторот на боја. Lab просторот на боја има три канали:

- L каналот ја означува осветленоста на сликата
- a каналот ја означува нијансата и сатурацијата на зелена-црвена боја
- b каналот ја означува нијансата и сатурацијата на сина-жолта боја.

Неговата идеја била таа да се земе L каналот од сликата како влез во невронската мрежа. Оттука, мрежата треба да научи да ги предвидува a и b каналите и да ги враќа истите како излез. Со комбинација на влезниот L канал од сликата, и предвидените ab канали, се добива финалната колоризирана слика.

Голема предност тука е тоа што за тренирање на моделот има огромен број на податоци кои што се истовремено и бесплатни: имено, било која слика може да се земе за тренирање бидејќи наголемиот дел слики достапни денес се во боја; потребно е само да се земе L каналот од сликата како влез и ab каналите како supervised сигнали. И претходни трудови за колоризација на слики се базираат врз тренирање на невронска мрежа, но со нивните

методи се добивале прилично незаситени (desaturated) колоризации. Во овој труд разликата е во справувањето со загубата (loss) на моделот.

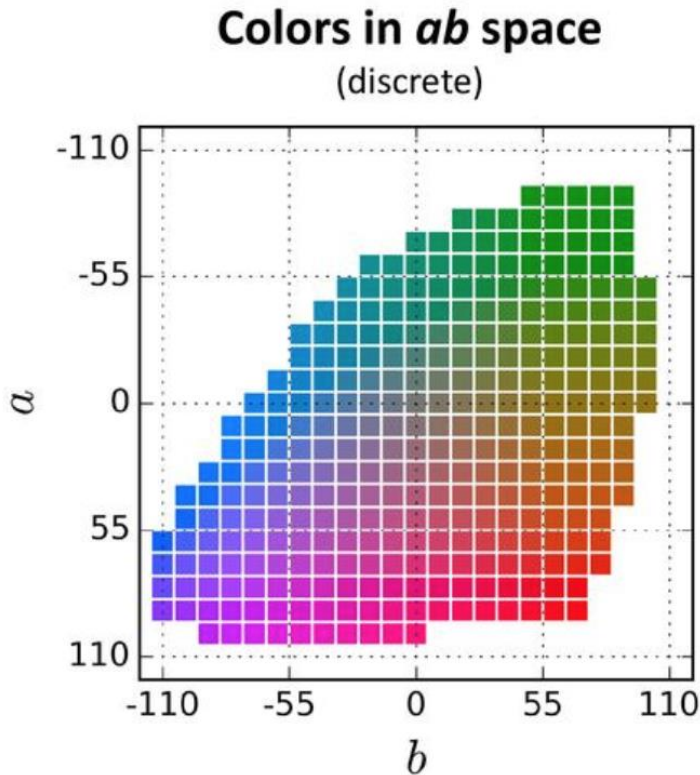
Предикцијата за боја е мултимодална – повеќе објекти може да имаат по неколку колоризации. На пример, јаболкото може да биде црвено, зелено и жолто, но тоа не може да биде сино. За соодветно да се моделира овој проблем, се прави предикција за дистрибуцијата на можни бои за секој пиксел. Потоа, се пресметува повторно загубата (loss) за време на тренирањето за да се потенцираат ретките бои. Ова му овозможува на моделот да ја искористи целосната разновидност на огромното количество на податоци на кои се тренира. На крај, се продуцира финалната колоризација со правење на mean annealing на дистрибуцијата. Крајниот резултат се колоризации кои се поживописни и пореалистични во однос на претходните пристапи кон оваа проблематика.

Дефинирање на проблемот со колоризација

Lab просторот на боја, исто како и RGB, е простор на боја определен со 3 канали, но за разлика од RGB просторот на боја, овде информацијата за бојата е енкодирана само во 2 канали: каналот a (зелена-црвена компонента) и каналот b (сина-жолта компонента). L (lightness) каналот ја претставува информацијата за интензитетот на светлината.

Grayscale сликата која сакаме да ја обоиме може да ја замислиме како L каналот од Lab просторот на боја и всушност целта е да се пронајдат a и b компонентите. Lab сликата која ќе се добие може да се трансформира во RGB формат со користење на стандардните трансформации за простори на боја (во OpenCV е тоа може да се направи во функцијата `cvtColor` со користење на опцијата `COLOR_BGR2Lab`).

За да се поедностават пресметките, ab просторот од Lab се квантизира со 313 бинови (bins). Наместо да се наоѓаат вредности за a и b за секој пиксел, поради ова квантизирање, едноставно се наоѓа број на bin од 0 до 312. Друг начин за дефинирање на овој проблем е тоа што ние веќе ги имаме вредностите од L каналот кои се во рангот од 0 до 255, па треба да ги најдеме ab каналите кои примаат вредности помеѓу 0 и 312. Со ова, задачата за предикција на боја станува мултиномен класификациски проблем каде за секој црно-бел пиксел има 313 класи од кои треба да одбереме.



Слика 2 - Квантизираните бои во *ab* просторот

Архитектура на Конволуциска Невронска Мрежа (CNN) за колоризација

Архитектурата предложена од Zhang et al е VGG тип на мрежа со неколку конволуциски блокови. VGG е конволуциска невронска мрежа за класификација и детекција, за прв пат претставена во трудот “Very Deep Convolutional Networks for Large-Scale Image Recognition”. Секој конволуциски блок има два или три конволуциски слоја по кои следува еден ReLU слој, за на крај да се заврши со Batch Normalization слој. За разлика од VGG мрежата, овде нема pooling или fully connected слоеви.

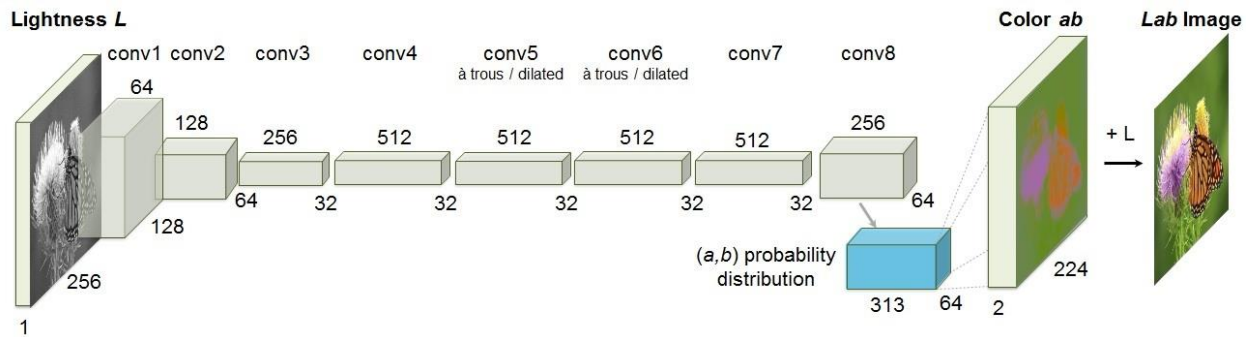
Влезната слика се скалира со димензии 224x224. Оваа скалирана влезна слика ако ја претставиме со X , кога ќе се прати во невронската мрежа дискутирана погоре, се трансформира во \hat{Z} од страна на мрежата. Математички, оваа трансформација се претставува како:

$$\hat{Z} = G(X)$$

Димензиите на \hat{Z} се $H \times W \times Q$, каде $H(= 56)$ и $W(= 56)$ се висината и ширината на излезот од последниот конволуциски слој. За секој од $H \times W$ пикселите, \hat{Z} содржи и вектор од $Q(= 313)$

Боење на црно-бели слики

вредности каде секоја вредност ја претставува веројатноста на секој пиксел кој што припаѓа на таа класа. Целта тука е да се најде еден пар на вредности за каналот ab за секоја веројатносна распределба $\hat{Z}_{h,w}$.



Слика 3 - CNN архитектура за моделот за колоризација

Добивање колоризирана слика од \hat{Z}

Конволуциската невронска мрежа прикажана на слика 3 ни дава колекција од дистрибуции во \hat{Z} од скалираната влезна слика X . Прашањето сега е како да добиеме еден пар на вредности за ab од секоја дистрибуција во \hat{Z} .

На прв поглед ова прашање изгледа лесно, едноставно може да се реши со земање на средната вредност на дистрибуцијата и избирање на ab парот кој што кореспондира на најблискиот квантизиран централен bin. За жал, ваквата дистрибуција не е Гаусова (нормална) распределба, и средната вредност од дистрибуцијата води кон неприродна, незаситена (desaturated) боја. На пример, бојата на небото е понекогаш сина, а понекогаш портокалово-розева. Распределбата на боите на небото е бимодална. Додека се обојува небото, и сината и жолтата боја ќе резултираат во задоволително добро обојување. Но, средната вредност на сина и жолта е некоја сива нијасна која не одговара за обојување на небо.

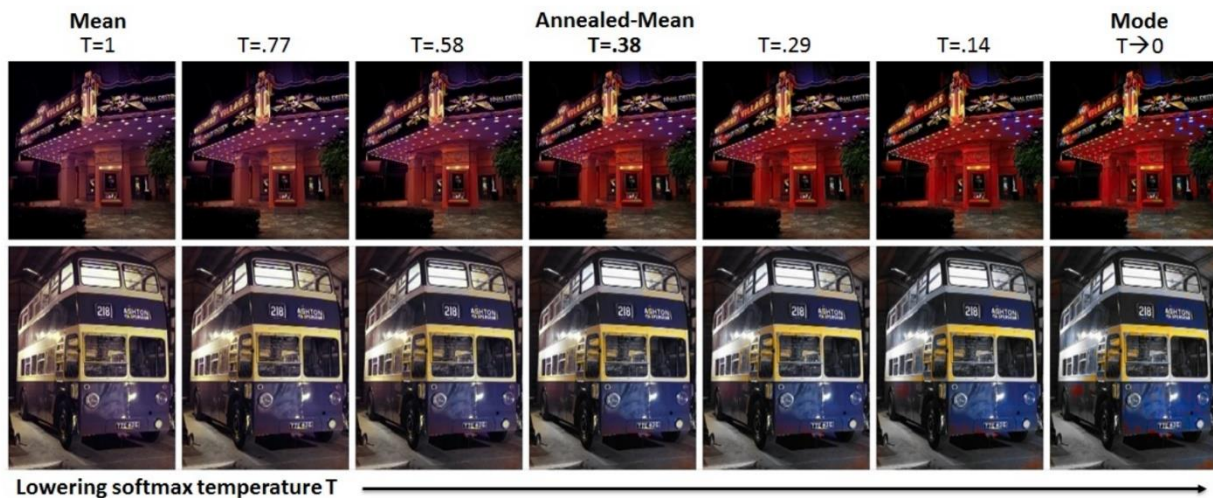
За решение на овој проблем би можело да се искористи модата на распределбата, така што наместо сиво, ќе се добие или сино или жолто небо. Но сепак, и покрај овој пристап со кој се добиваат поживописни бои, понекогаш се расипува просторната конзистентност. Решението на авторите на овој труд било да се интерполира помеѓу средната вредност и модата за да се добие вредност која што се нарекува annealed-mean. Параметар наречен температура (T) бил искористен за да се контролира степенот на интерполација. Крајна вредност за $T=0.38$ се искористила како trade-off помеѓу двете екстреми.

Боење на црно-бели слики

Парот \hat{a}_b којшто кореспондира на annealed-mean средината на дистрибуцијата $\hat{Z}_{h,w}$ е претставен во $\hat{Y}_{h,w}$, којшто може да биде запишан како трансформација на оригиналната дистрибуција $\hat{Z}_{h,w}$

$$\hat{Y} = H(\hat{Z})$$

Како што сликата минува низ невронската мрежа, нејзината големина се намалува на 56x56. Поради тоа предвидената \hat{a}_b слика, \hat{Y} , исто така има димензија 56x56. За да се добие колоризираната слика, предвидената \hat{a}_b слика се конвертира во оригиналната димензија на сликата и потоа се додава каналот за светлината, L , за да се добие финалната слика во боја.



Слика 4 - Annealed-mean, заедно со температурата (T), се користи за интерполација помеѓу средната вредност и модата на распределбата

Мултиномна функција на загуба со ребаланс на боја

Сите невронски мрежи се тренираат со дефинирање на некоја функција на загуба (Loss function). Целта на процесот на тренирање е да се минимизира загубата на тренирачкиот сет. Во проблемот за колоризација, податоците за тренирање се содржат од илјадници слики во боја и нивните црно-бели варијанти.

Излезот од конволуциската невронска мрежа за дадена влезна слика X е \hat{Z} . Ние треба да ги трансформираме сите слики во боја од тренирачкиот сет во нивните соодветни Z вредности. Математички, сакаме да го инвертираме мапингот H

$$Z = H^{-1}(Y)$$

За секој пиксел, $Y_{h,w}$, од некоја излезна слика Y , можеме да го најдеме најблискиот аб бин и да го претставиме $Z_{h,w}$ како one-hot вектор, во кој ќе доделиме вредност 1 за најблискиот аб бин и 0 за останатите 312 бинови. Но за подобар резултат, најблиските 5 соседи се земаат во предвид и се користи Гаусова (Нормална) распределба за пресметување на дистрибуцијата на $Z_{h,w}$ врз основа на растојанието од вистинската вредност.

Во други случаи на Конволуциски невронски мрежи, кога станува збор за класификација се користи стандардната cross-entropy како функција на загуба за да се споредат вистинската вредност Z и предвидената \hat{Z} со користење на:

$$L(\hat{Z}, Z) = - \frac{1}{HW} \sum_{h,w} \sum_q Z_{h,w,q} \log(\hat{Z}_{h,w,q})$$

Оваа функција на загуба, за жал, продуцира доста бледи бои. Тоа е така поради тоа што распределбата на боите во ImageNet датасетот е наклонет околу линијата на сива боја.

Ребаланс на боја

За да се поттикне алгоритмот да продуцира поживи бои, авторите на овој труд решиле да ја сменат функцијата на загуба во:

$$L(\hat{Z}, Z) = - \frac{1}{HW} \sum_{h,w} v(Z_{h,w}) \sum_q Z_{h,w,q} \log(\hat{Z}_{h,w,q})$$

Членот во равенката за ребаланс на боја, $v(\cdot)$, се користи за да се ребалансира загубата која се случува поради реткоста на таа класа на боја. Ова придонесува кон добивање бои кои се поживи и позаситени како резултат од невронската мрежа.

Резултат од колоризацијата

Авторите на овој труд имаат истренирано и објавено два Caffe модели – со и без ребалансирање на бојата. На сликите подолу се прикажани резултати од двете верзии на моделите. Средната колона ја прикажува верзијата без ребаланс а последната колона го прикажува резултатот од моделот со ребалансирање на бојата.

Како што можеме да видиме, ребалансот на бојата ни дава како резултат слики кои што се со доста поинтензивни бои и поживописни. Повеќето од нив се бои коишто изгледаат веродостојно во колоризацијата. Од друга страна пак, понекогаш овој модел може да даде делови каде што боите ќе се премногу заситени и интензивни и ќе има поголем контраст, што на човечкото око тоа му изгледа доста неприродно.

Боење на црно-бели слики

Кога конвертираме црно-бела слика, може да има повеќе веродостојни верзии во боја. За да евалуираме дали е добра колоризацијата не треба да споредуваме со тоа колку добро точно се совпаѓа со вредностите на оригиналната слика во боја, туку колку сликата му изгледа веростојно и пријатно на човечкото око.



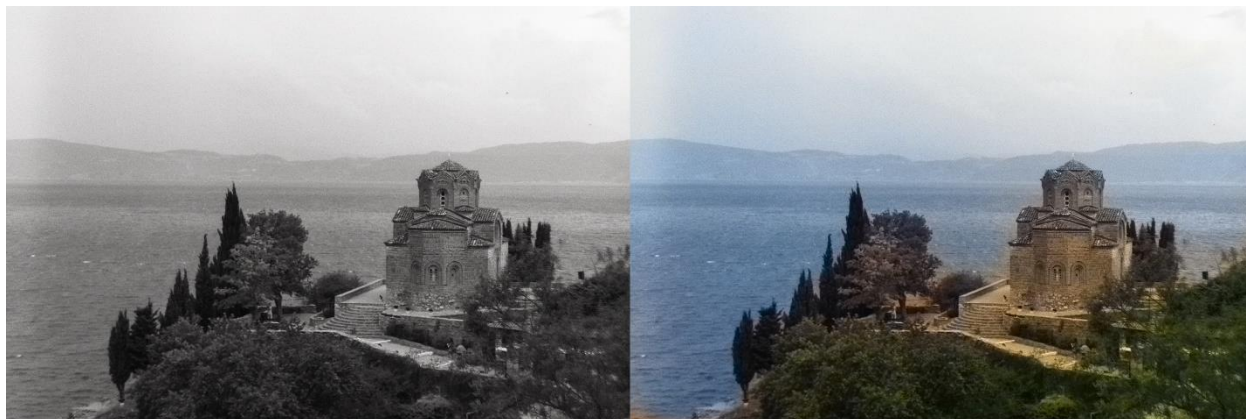
Колоризирање на оригинал црно-бели слики

За тренирање на моделот авторите на овој труд искористиле „лажни“ црно-бели слики, односно слики кои биле во боја па од нив биле извлечени и исфрлени ab каналите. Но сепак моделот продуцира добри колоризации дури и на стари фотографии кои биле оригинално фотографирани во црно-бел формат, дури и покрај тоа што ниските статистики за оригинал црно-белите слики се доста различни од статистиките на оние слики од модерното време на кои моделот бил трениран.



Слика 5 - Стара слика од стариот дел на Охрид, улица и куќи со препознатливата Охридска архитектура

Боење на црно-бели слики



Слика 6 - Црквата св. Јован Канео во Охрид



Слика 7 - Стара слика од манастирот св. Наум Охридски на брегот на Охридското езеро



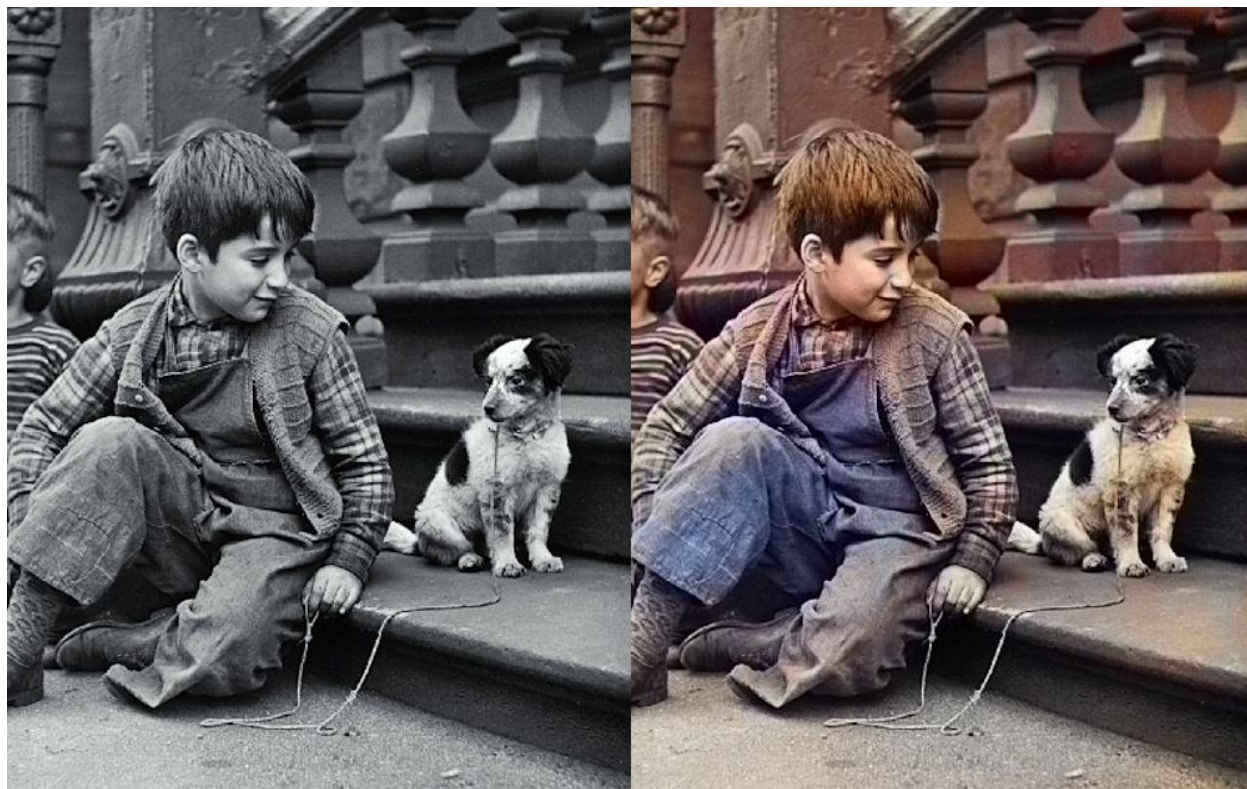
Слика 8 - Пејзаж од стариот дел на градот Охрид заедно со црквата св. Јован Канео

Боење на црно-бели слики



Слика 9 - Адолф Хитлер

Други примери за колоризација



Боење на црно-бели слики



Заклучок

Покрај тоа што колоризација на црно-бели слики е задача од областа на компјутерска графика, тоа е исто така и пример за сложен проблем на предвидување на пиксели во компјутерска визија. Тука се покажа дека колоризацијата со длабока конволуциска невронска мрежа и добро-одбрана функција на загуба се приближува кон продуцирање на резултати кои не се разликуваат од вистинските слики во боја. Методот опишан овде не само што обезбедува корисен графички излез, туку исто така може да биде гледан како pretext task за репрезентативно учење, користен во само-надгледувано учење (self-supervised learning). Покрај тоа што е тренирана на боја, нашата мрежа учи и репрезентации коишто се корисни за класификација на објекти, детекција и сегментација, и тоа со добра точност и прецизност споредувано со други self-supervised претренирачки методи.

Референци

1. Richard Zhang, Phillip Isola, Alexei A. Efros – Colorful Image Colorization, European Conference on Computer Vision (ECCV) 2016
2. Convolutional Neural Network based Image Colorization using OpenCV - <https://learnopencv.com/convolutional-neural-network-based-image-colorization-using-opencv/>
3. Black and white image colorization with OpenCV and Deep Learning - <https://www.pyimagesearch.com/2019/02/25/black-and-white-image-colorization-with-opencv-and-deep-learning/>
4. Github repo of richzhang: <https://github.com/richzhang/colorization>
5. Deep Learning Project – Colorize Black & White Images with Python - <https://techvidvan.com/tutorials/deep-learning-project-colorize-black-white-images-with-python/>
6. Image Colorization using Convolutional Autoencoders - <https://towardsdatascience.com/image-colorization-using-convolutional-autoencoders-fdabc1cb1dbe>
7. Deep Learning based image colorization with OpenCV - <https://cv-tricks.com/opencv/deep-learning-image-colorization/>