Abstract:

- fully automated

- demonstrating the viability of our methodology and revealing promising avenues for future work.

1. Introduction

- el automated colorization has been subject to much research within the computer vision and machine learning communities

- el AI we el ML we el ttwr da 3ml outputs gameela fe kza 7aga zy e3adt el video we ta7sen el image

- est5dmo statistical-learning-driven approach 3shan ye7llo el problem de

- el output bytl3 bona2n 3la images elly et3ml mnha faqqat (el dataset) mn 8er ay user interaction

- fe el senen el a5era el CNN zhr fe el image classification we 7a22a2 error rates lower than 4% in the ImageNet challenge

- sabb naga7 CNN hwa qodreto 3la tamyez el colors, el patterns we el shapes elly fe el images and associate them with object classes.

- homa by3tqdo b2a en el characteristics dol tabe3y enhom ye5llo nafsohom by3rfo yelwno el images

la22n el object classes, patterns, and shapes generally loh 3elaqa be e5tyar el lon

2. Related work

- el project bta3hom ispired in part wa7ed tany CNN based we automatically colorizing images bardo, el tany da mo3tmd 3la several ImageNet-trained layers from VGG16

-* VGG16 (also called OxfordNet) is a convolutional neural network architecture named after the Visual Geometry Group from Oxford,

who developed it. It was used to win the ILSVR (ImageNet) competition in 2014.

To this day is it still considered to be an excellent vision model,

although it has been somewhat outperformed by more revent advances such as Inception and ResNet.

3. Approach

- bybno learning pipeline bttkwn mn neural network and an image pre-processing front-end

3.1. General pipeline

- fe el training time by2ro images 224 x 224 and 3 channels lel RGB color space

el images de btt7wl le CIELUV color space

The black and white luminance L channel is fed to the model as input.

The U and V channels are extracted as the target values

-* LUV --> Luminance, U and V channels

-* el U we el V dol byb2o shaylen el color information we el L el eda2a

-* In colorimetry, CIELUV is a color space that is extensively used for applications such as computer graphics which deal with colored lights

-* Colorimetry is "the science and technology used to quantify and describe physically the human color perception."

- fe el test time el input byb2a 224 x 224 x 1 b&w image

byt3ml 2 arrays wa7ed lel U channel we wa7ed lel V channel kol array el size bta3o 224 x 224 x 1

keda b2a m3ana 3 arrays wa7ed lel U we wa7ed lel V we wa7ed mn el awl ka input lel L el tlata arrays dol b2a byt-concatenated together ye3mlolna el predicted image

3.2. Transfer learning

"We initialized parts of model with a VGG16 instance that has been pretrained on the ImageNet dataset. Since image subject matter often implies color palette, we reason that a network that has demonstrated prowess in discriminating amongst the many classes present in the ImageNet dataset would serve well as the basis for our network. This motivates our decision to apply transfer learning in this manner."

3.3. Activation function

- Rectified -> $f(x) = max(0; x)$
- The rectified linear unit has been empirically shown to greatly accelerate training convergence
- feh 3eb fe est5dam Rectified fun en fe 7alat dayman hatkon be 0

3.4. Batch normalization

- Ioffe et al

3.5. Baseline regression model

- We describe this architecture as comprising a “summarizing”, encoding process on the left side followed by a “creating”, decoding process on the right side

3.6. Final classification model

- The regression model suffers from a dimming problem because it minimizes some variant of the Lp norm

- el suffer da b2a motivates the model to choose an average or intermediate color when multiple distinct color choices are possible

- 8ayro model el problem le classification problem

- In order to perform classification on continuous data, we must discretize the domain

- el U we el V bto3 el CIELUV color space kano bya5do values fe interval [-100, 100] hay5lhom ya5do values fe interval [0,49]

- el discreting da 7sl by applying a binning function (denoted bin()) to each input image prior to feeding it to the input of the network

- this function returns an array of the same shape as the original image with each U and V value (elly heya mn 0 le 49)

- Then, instead of directly predicting numeric values for U and V , the network outputs two separate sets of the most probable bin

numbers for the pixels, one for each channel. We used the sum of cross-entropy loss on the two channels as our minimization objective

- Combining multiple intermediate feature maps in this fashion has been shown to increase prediction quality in segmentation problems,

producing finer details and cleaner edges. Although there is no explicit segmentation step in our setup, this approximate approach allows our system to

minimize the amount of visual noise that is generated along object edges in the output image

- fe el a5er, the concatenation layer is followed by three 3 x 3 convolutional layers

which are in turn followed by the final two parallel 1 x 1 convolutional layers corresponding to the U and V channels

- These 1 1 convolutional layers act as the fully-connected layers to produce 50 class scores for each channel for each pixel of the image

the classes with the largest scores on each channel are then selected as the predicted bin numbers via an un-binning function, we then convert the predicted bins back to numerical U and V values using the means of the selected bins.

4. Dataset

- MIT CVCL Urban, Natural Scene Categories -> dataset feha alaaf el images byt2smo le 8 categories

- homa garbo 411 images mn category "Open Country"

- garrbo 3la The McGill Calibrated Colour Image Database bardo (feha alaf el natural scenes organized by categories) homa garrbo 3la sample mn kol category

- grabo 3la classes "spatula, school bus, bear,book shelf, armor, kangaroo, spider, sweater, hair dryer, and bird" mn ILSVRC 2015 CLS-LOC

-* The ILSVRC 2015 CLS-LOC dataset is the dataset used for the ImageNet challenge in 2015

- The MIRFLICKR dataset comprises 25000 Creative Commons images downloaded from the community photosharing website (feha categories keter gedn)

- 3mlo scale le kol image le 224 x 224 x 3 , we 3mlo nos5a grayscale b dimention 224 x 224 x 1

- Since the input of our network is the input of the ImageNet-trained VGG16, which expects its input images to be zero-centered and of

dimensions 224 x 224 x 3, we duplicate the grayscale image three times to form a (224 224 3)-sized image and subtract the mean R, G, and B value

across all the pictures in the ImageNet dataset. The resulting final image serves as the black-and-white input image for the network.

5. Experiments

- fe el regression el closeness 7sboh b el fr2 maben el U we el V bta3 el generated image we el actual image (L = |Ud - Ut|^2 + |Vd - Vt|^2)

- fe el classification el closeness 7sboh b nesbt el binned pixel values elly bt-match maben el generated wel actual image le channel U & V

- Acc.U = 1/N^2 * sum (equal binned pixel values le channel U)   nested loop 3l pixels we condition

- Acc.V = 1/N^2 * sum (equal binned pixel values le channel V)

- Performances on MIT CVCL Open Country test set:

System	Acc. (U)	Acc. (V)	Sat. diff.
Classification	34.64%	24.14%	6.5%
Regression	12.92%	19.02%	85.8%

6. Conclusion and future work

- asbto el efficacy and potential of using deep convolutional neural networks to colorize black and white images

- asbto en e3tbar el problem classfication a7sn bkter mn enha teb2a baseline regression (and thus shows much promise for further development)

- Finally, redesigning the system around an adversarial network may yield improved results,

since instead of focusing on minimizing the cross-entropy loss on a perpixel basis,

the system would learn to generate pictures that compare well with real-world images

Based on the quality of results we have produced, the network we have designed and built would be a prime candidate for being the generator

in such an adversarial network.

References

- 7atten 14 references