Automatic colorization of natural images using deep learning

Sharique Alam khan[1], Alok Katiyar[2]
[1,2](Students of BCA (SCSE) Galgotias University, Greater Noida, Uttar Pardesh)

Abstract

An approach based on deep learning for automatic colorization of image with optional user-guided hints. The system maps a grey-scale image, along with, user hints" (selected colors) to an output colorization with a Convolution Neural Network (CNN). Previous approaches have relied heavily on user input which results in non-real-time desaturated outputs. The network takes user edits by fusing low-level information of source with high-level information, learned from large-scale data. Some networks are trained on a large data

set to eliminate this dependency. The image colorization systems find their applications in astronomical photography, CCTV footage, electron microscopy, etc. The various approaches combine color data from large data sets and user inputs provide a model for accurate and efficient colorization of grey-scale images. Keywords—image colorization; deep learning; convolutional neural network; image processing.

Introduction

There are two broad approaches followed in computer graphics to image colorization: user-driven edit propagation and automatic colorization driven by data. In the first approach, a user provides hints by colored strokes over a grey-scale image. The enhancement strategy at that point produces a colorized picture it coordinates the client's strokes, while additionally sticking to hand characterized picture priors, for example, piece-wise smoothness. These strategies can be utilized to yield extremely noteworthy outcomes however they regularly require serious client collaboration (once in a while in excess of fifty strokes), as each diversely hued picture area must be unequivocally shown by the client. Since the framework depends intensely on client contributions for hues, even locales with little shading vulnerability, for example, blue sky or green vegetation, should be determined. An additional information driven colorization strategy is like- wise attempted by the scientists to address these constraints. These techniques colorize a dim scale picture in one of two different ways: either by coordinating it with a model hued picture in a database and non-parametric duplicating hues from that photograph, or by taking in parametric mappings from dark scale to shading from substantial scale picture information. The latest techniques in information driven worldview pro- posed, utilize profound neural systems and they are completely programmed. The aftereffects of the procedure more often than not contain incorrect hues and ancient rarities alongside the shaded yield. The precise shade of couple of nonexclusive articles, for example, a shirt, is regularly vague it's shading could be orange, red, or pink. The new methodology is to attempting to consolidate both of these strategies to outdo both, utilizing substantial scale information to

learn priors about regular shading symbolism, while in the meantime joining client decisions. The thought is to prepare a CNN on an extensive informational collection to straight forwardly outline scale pictures, alongside client con- tributions, to create a colorized yield picture. Amid preparing, haphazardly recreating client inputs empowers us to counter the issue of gathering an exceptionally substantial number of client cooperations. Colorization of grayscale images is a simple task for the human imagination. A human need only recall that sky is blue and grass is green; for many objects, the mind is free to hallucinate several plausible colors. The high-level comprehension required for this process is precisely why the development of fully automatic colorization algorithms remains a challenge. Colorization is thus intriguing beyond its immediate practical utility in graphics applications. Automatic colorization serves as a proxy measure for visual understanding. Our work makes this connection explicit; we unify a colorization pipeline with the type of deep neural architectures driving advances in image classification and object detection.

Literature Reviews

Colorization basically involves assigning colors to grey-scale images. Convolutional neural networks are basically designed to deal with image data. Many authors have done promising work on this idea.

- Domonkos varga proposed the idea of automatic coloring of cartoons images, since they are very different from natural images, they pose a difficulty as their colors depend on artist to artist. So, the data set was specifically trained for cartoon images about 100000 images, 70% of which were used in training and rest for validation. But unfortunately, the colour uncertainity is much higher than in natural images and evaluation is subjective and slow.
- Shweta slave proposed another similar approach, employing the use of google's image classifier. The system model is divided into 4 parts **encoder , feature extractor , fusion layer and decoder.** The system is able to produce acceptable outputs , given enough resources , CPU , memory ,and large data set. This is mainly proof of concept implementation .
- Yu chen proposed a approach to mainly address the problem of coloring Chinese films from the past. They used existing data set with their data set of Chinese images, fine- tuning the overall model. The network makes use of multi-scale convolutional kernels, combining low and middle features.
- V.K. Putri proposed a method to convert plain sketches into colorful images. It uses sketch inversion model and color prediction in CIE Lab color space. This approach is able to handle handdrawn sketches including various geometric transformations. The limitation found was that, data-set is very limited but it works well for uncontrolled conditions.
- Richard Zhang has proposed a optimized solution by using huge data-set and single feed-forward pass in CNN. Their main focus lies on training part. They used human subjects to test the results and were able to fool 32% of them. can have various number of neurons. The various attempts used various architectures . In some papers, generally number of neurons is same as the dimension of the feature descriptor extracted from each pixel coordinates in a grey-scale image.

2) Global features : Most of the methods utilize the global features to form an image filter, and then use this filter to select similar images from a large image set automatically. However, some models produced unnatural colorization result due to global similarity but semantic difference.

3) Data set size: Parametric and Non- Parametric models use different sizes of data sets for training the CNN. The Parametric model use very large data set to to train the CNN and produce more accurate result while the Non-Parametric models rely more on the input hints and reference image and use smaller data set for training the CNN.

4) Semantic information: Semantic information has a sig- nificant and unavoidable role in deep image colorization. For effective colorization of images, the system must have information of the semantic composition of the image and its localization. For instance, leaves on a tree may be colored some kind of green in spring, but they should be colored brown for a scene set in autumn. VGG-16 CNN model was used by majority of approaches to extract semantic information about the image before applying colorization techniques.

5) Feature extraction: Features of the image are obtained through by integrating pre-trained neural networks to extract information about objects, shapes and use this context to assign color values to the objects. Some approaches used using Inception ResNet V2 classifier or Tensorflow to serve this purpose.

6) Surveyed techniques: The methods for image colorization can be categorized into two major groups: Based on user inputs and automatic colorization based. The methods make use of CNN for the colorization. The non parametric methods first define one or more color reference images using the input from either user or a source image as source data. Then, color is transferred onto the input image from matching regions of the reference data. On the other hand, Parametric methods learn from training on large data sets of colored images, using either regression onto continuous color space or classification of quantized color values.

7) Survey details (architecture) : The deep neural networks consists of an input layer, many hidden layers and an output layer.

RELATED WORK

- Previous colorization methods broadly fall into three categories: scribble-based , transfer and automatic direct prediction
- *Scribble-based* methods, introduced by Levin *et al.* require manually specifying desired colors of certain regions. These scribble colors are propagated under the assumption that adjacent pixels with similar luminance should have similar color, with the optimization relying on Normalized Cuts Users can interactively refine results via additional scribbles. Further advances extend similarity to texture and exploit edges to reduce color bleeding .
- *Transfer-based* methods rely on availability of related *reference* image(s), from which color is transferred to the target grayscale image. Mapping between source and target is established automatically, using correspondences between local descriptors , or in combination with manual intervention . Excepting , reference image selection is at least partially manual.
- In contrast to these method families, our goal is *fully automatic* colorization. We are aware of two recent efforts in this direction. Deshpande *et al.* colorize an entire image by solving a linear

system. This can be seen as an extension of patch-matching techniques adding interaction terms for spatial consistency. Regression trees address the high-dimensionality of the system. Inference requires an iterative algorithm. Most of the experiments are focused on a dataset (SUN-6) limited to images of a few scene classes, and best results are obtained when the scene class is known at test time. They also examine another partially automatic task, in which a desired global color histogram is provided.

- The work of Cheng *et al.* is perhaps most related to ours. It combines three levels of features with increasing receptive field: the raw image patch, DAISY features , and semantic features . These features are concatenated and fed into a three-layer fully connected neural network trained with an L2L2 loss. Only this last component is optimized; the feature representations are fixed.
- Unlike , our system does not rely on hand-crafted features, is trained end-to-end, and treats color prediction as a histogram estimation task rather than as regression. Experiments in Sect. justify these principles by demonstrating performance superior to the best reported by across all regimes.
- Two concurrent efforts also present feed-forward networks trained end-to-end for colorization. Iizuka *et al.*propose a network that concatenates two separate paths, specializing in global and local features, respectively. This concatenation can be seen as a two-tiered hyper column; in comparison, our 16-layer hyper column creates a continuum between low- and high-level features. Their network is trained jointly for classification (cross-entropy) and colorization (L2L2 loss in Lab). We initialize, but do not anchor, our system to a classification-based network, allowing for fine-tuning of colorization on unlabelled datasets.
- Zhang *et al.* similarly propose predicting color histograms to handle multi-modality. Some key differences include their usage of up-convolutional layers, deep supervision, and dense training. In comparison, we use a fully convolutional approach, with deep supervision implicit in the hyper column design, and, as Sect. 3 describes, memory-efficient training via spatially sparse samples.

Working of project

We started by trying to overfit our model on a 270-image random subset of ImageNet data.

To determine a suitable learning rate, we ran multiple trials of training with minibatch updates to see which learning rate yielded faster convergence behavior over a fixed number of iterations. Within the set of learning rates sampled on a logarithmic scale, we found that a learning rate of 0.001 achieved one of the largest per-iteration decreases in training loss as well as the lowest training loss of the learning rates sampled. Using that as a starting point, we moved to with the entire training set. With a hold-out proportion of 10% as the validation set, we observed fastest convergence with a learning rate of 0.0003

We also experimented with different update rules, namely Adam [8] and Nesterov momentum [14]. We followed the recommended $\beta_1 = 0.9$ and $\beta_2 = 0.99,0.999$. For Nesterov Momentum, we used a momentum of 0.9. Among these options, the Adam update rule with $\beta_1 = 0.9$ and $\beta_2 = 0.999$ produced slightly faster convergence than the others, so we used the Adam update rule with these hyperparameters for our final model.

Among these options, the Adam update rule with $\beta_1 = 0.9$ and $\beta_2 = 0.999$ produced slightly faster convergence than the others, so we used the Adam update rule with these hyperparameters for our final model.

In terms of minibatch sizes, we experimented with batches of four, six, eight and twelve images based on network architecture. Some alternative structures we tried required less memory usage, so we tested those with all four options. The model shown in Figure 3, however, is memory-intensive. Due to the limited access of computational resource, we were only able to test it with batch sizes of four and six with the GPU instance. Nevertheless, this model with a batch size of six demonstrated faster and stabler convergence than the other combinations.

For weight initialization, since our model uses the rectified linear unit as its activation function, we followed the Xavier Initialization scheme proposed by [3] for our original trainable layers in the decoding, "creating" phase of the network.

We also developed several alternative network structures before we arrived at our final classification model. The following are some design elements and decisions we weighed:

Multilayer aggregation – elementwise sum versus concatenation: we experimented with performing layer aggregation using an elementwise sum layer in place of the concatenation layer. An elementwise sum layer reduces memory usage, but in our experiments, it turned out to harm training and prediction performance.

Presence or absence of residual encoder units: a residual encoder unit refers to a joint convolution elementwise-sum step on a feature map in the "summarizing" process and an upscaled feature map in the "creating" process, as described in Section 3. We experimented with trimming away the residual encoder units and applying aggregation layers directly on top of the max pooling layers inherited from VGG16. However, the capacity of the resulting model is much smaller, and it showed poorer quality of results when overfitting to the 300-image subset.

The final sequence of convolutional layers before the network output: we experimented with one and two convolutional layers with various depths, but the three-layer structure with the current choice of depths yielded the best results.

Color space: initially, we experimented with the HSV color space to address the under-saturation problem. In HSV, saturation is explicitly modeled as the individual S channel. Unfortunately, the results were not satisfying. Its main issue lies in its exact potential merit: since saturation is directly estimated by the model, any prediction error became extremely noticeable, making the images noisy.

Results and Conclusion

We have presented a method of fully automatic colorization of unique greyscale images combining state of-the-art CNN techniques. Using color representation and the right loss function, we have represented that the method is capable of producing a plausible and vibrant colorization of certain parts of images. Our model does very well with the animals like cats and dogs because the dataset we chose i.e., ImageNet consists large amount of pictures of these animals [12]. Even the outdoor scenes turnout very good with our model. The model also captures notion of sunset and paints it orange. The model produces plausible images even with the sketches. Through our experiments, we have demonstrated the efficacy and potential of

using deep convolutional neural networks to colorize black and white images. In particular, we have empirically shown that formulating the task as a classification problem can yield colorized images that are arguably much more aesthetically-pleasing than those generated by a baseline regression-based model, and thus shows much promise for further development.

Our work therefore lays a solid foundation for future work. Moving forward, we have identified several avenues for improving our current system. To address the issue of color inconsistency, we can consider incorporating segmentation to enforce uniformity in color within segments. We can also utilize post-processing schemes such as total variation minimization and conditional random fields to achieve a similar end. 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

1. Z., Yang, Q., Sheng, B.: Deep colorization. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 415–423 (2015)Google Scholar.

2. Dahl, R.: Automatic colorization (2016). http://tinyclouds.org/colorize/.

3. Charpiat, G., Hofmann, M., Schölkopf, B.: Automatic image colorization via multimodal predictions. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) ECCV 2008, Part III. LNCS, vol. 5304, pp. 126–139. Springer, Heidelberg (2008)CrossRefGoogle Scholar.

4. Ramanarayanan, G., Ferwerda, J., Walter, B., Bala, K.: Visual equivalence: towards a new standard for image fidelity. ACM Trans. Graph. (TOG) **26**(3), 76 (2007)CrossRefGoogle Scholar.

5. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014).

6. Bengio, Y., Courville, A., Vincent, P.: Representation learning: a review and new perspectives. IEEE Trans. Pattern Anal. Mach. Intell. **35**(8), 1798–1828 (2013)CrossRefGoogle Scholar.

7. Ngiam, J., Khosla, A., Kim, M., Nam, J., Lee, H., Ng, A.Y.: Multimodal deep learning. In: Proceedings of the 28th International Conference on Machine Learning (ICML 2011), pp. 689–696 (2011)Google Scholar.