Convolutional Neural Network-based Image Colorization

Department of Computer science and Engineering
Amrita Vishwa Vidyapeetham
Amritapuri Campus
Kollam 690525
Kerala

Namitha S, Ritika R prasad, Lakshmi G Pillai

Mathematics for Intelligent Systems 3-19MAT204
S3 B.Tech CSE(AI)

***Abstract*—In this paper, we tend to point out colourizing grey-scale pictures victimization deep learning ways. We propose a completely automatic approach that produces beamy and realistic colourizations. The system is materialized victimization CNNs as they're used for image classification and recognition owing to its high accuracy. The CNN network used will be pre-trained on 1.m images from imagenet. This approach leads to mod and progressive performance on many feature learning benchmarks. The algorithmic rule project contains the methodology for automatic image colourization supported feature extraction like convert to science laboratory area, size to 256x256, colourize, and concatenate to the first full resolution, and convert to RGB.**

Keywords—CNN, Imagenet, Lab colour space ,rgb color space

I. INTRODUCTION

Grayscale image colourization could be a new image process topic, however, totally different trails for manual grayscale colourization were found within the 80s. The colourization mechanism leads several researchers throughout a previous couple of years to seek out an additional application from this technology not solely giving colours to uncoloured pictures but additionally eliminating colours from the colour pictures and videos and recolouring them back to build profit from black and white pictures and videos options. Our main goal is to provide an inexpensive image colourization from grayscale pictures. Considering the grayscale image in (FIG5) a, apparently initially visible it appears to be a monochrome image created from black and white image, and the entire scales of reminder grey. However, we have a tendency to don't seem to be aiming at gaining the particular ground truth colour. Our task is to provide chromatic entrancing results. The rule we have a tendency to use here relies on Convolutional neural networks(CNN). A Convolutional Neural Network (ConvNet/CNN) could be a Deep Learning rule which might soak up an associate in the nursing input image, designate significance to numerous aspects/objects within the image and be ready to designate one from the opposite.CNN's are preponderantly used for image classification and recognition. The speciality of CNN is its convolutional ability. The potential for additional uses of CNNs is limitless and desires to be investigated and pushed to additional boundaries to find all that may be achieved by this tangled machinery.

What we are trying to do is predict the distribution of possible colours for each input image pixel. At training time, we weight the loss to highlight unusual colours. This helps to allow our model to take advantage of the full diversity of the large-scale data it is trained on. Finally, by taking the annealed distribution average, we generate a final colourization.

The end result is more vivid and perceptually realistic colourizations than those of previous methods.

II. LITERATURE REVIEW

Zhang, R., Isola, P. and Efros, A.A., 2016. Colourful image colourization which was published with a paper titled Colorful Image Colorization during which they presented a Convolutional Neural Network for colourizing grey images was taken as a baseline of the project. We've gotten their model and tested the pre-trained model. Within the model, given a grayscale photograph as input, brings up the concept of hallucinating a plausible colour version of the photograph.

It proposes a wholly automatic approach that produces vibrant and world colourizations. The system is implemented as a feed-forward pass during a CNN at test time and is trained on over 1,000,000 colour images. They need also evaluated their algorithm employing a "colourization Turing test," asking human participants to determine between a coloured image output and ground truth colour image and also the tactic successfully fools humans on 32% of the trials.
Another paper which was referred to was published by Zezhou Cheng, Qingxiong Yang, Bin Sheng introduces a new colourization, fully automatic, Method of reducing consumer effort using deep neural networks and the reliance on the colour images of the example. Informative yet discriminatory characteristics, including patch function, a new symbolic feature and DAISY function are extracted and extracted. It acts as the neural network's data. To ensure artefact-free colourization efficiency, the output chrominance values are further refined using joint bilateral filtering. Numerous experiments show that our solution outperforms in terms of both quality and speed, state-of-the-art algorithms.

III. Approach

In our method, we have prediction functions from wide colour image datasets at training time, proposing the problem as a regression on continuous colour space model information to classify colours with a larger model, trained on more data, and with several loss function innovations and mapping to final continuous output. The trained data is built by initially designing an appropriate objective function that handles the multimodal uncertainty of the colourization problem and captures a wide variety of colours, then by training on a million colour photos. We classify colours, with a larger model, trained on more data, and with several innovations in the loss function and mapping to final continuous output.

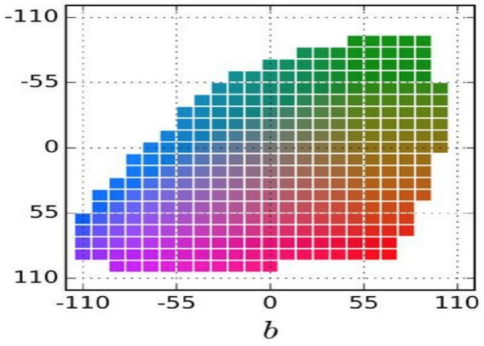
Objective

Our approach consists of coaching a CNN with coloured images from a large dataset allowing it to be told colour features with none human supervision. We pass the L (lightness) channel because the input to our CNN which further outputs the ab channels adore the colours for the grayscale image and so concatenate the input and output to come up with the total 3 channel images. We then convert this back to the RGB colour space. The remainder of our paper is discussing the main points of coaching and implementation together with the result. We are defining our colourization problem in regard to the CIELab colour space. Also said as L a b. The CIELAB colour space was established in 1976 by the International Commission on Illumination.

Colour is portrayed as three values: L for perceptual lightness and a and b* for biological vision's 4 main colours: red, green, blue and yellow. Within the Lab colour space, the grayscale image we wish to paint is often considered because of the L-channel of the image and our intention is to search out the a and b components. Using generic colour space transformations, the Lab image thus obtained is often mapped to the RGB colour space. The ab region of the Lab colour space is principally measured into 313 bins to simplify the calculations. We are finding a bin number between 0 and 312. This means that we have already got the L channel that takes values from 0 to 255 and then we have to find the ab channel that takes values between 0 and 312. Therefore, the role of colour prediction is now becoming a multinomial classification problem where there are 313 groups to decide on from for each grey pixel.

Lab colour space

A lab colour space consists of mainly 3 parts. L a and b. L is the luminance, a represents the values from green to magenta and b represents the colours from blue to yellow. Here 0 values of L represents the absence of luminance and higher values represent the presence of the luminance. Lower values of the channel represent the presence of green and higher values represent the magenta colour. Similarly in b channel, lower represent blue and higher represent yellow lab colour space is used for the image colourization of our project because this space minimizes the effect of correlation between the colour channels. The ab colour space is shown in (FIG1).



(FIG1): Colour in ab space

RGB colour space

It is designed with a perspective of the human sense of colour perceptions. This colour space represents each pixel with three values of Red, Green and Blue. Each of these pixels on the screen represents 3 light-emitting devices which emit R, G, and B to represent the colour of that pixel. When each of the

colours is represented by bytes, we can have 256 shades of each colour which will result in having a variety of colours with 26 shades from each other. This colour space is mainly designed to represent colours with electromagnetic systems, like computers, televisions, printers and so on.

Convolution neural network

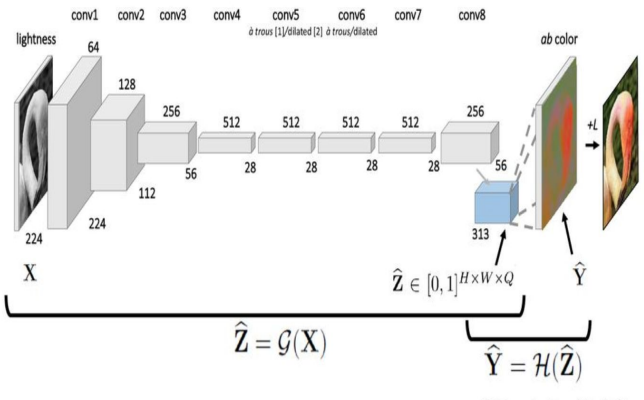
Convolutional neural network (CNN, or ConvNet) is a category of deep neural networks in the field of deep learning. A convolutional neural network is pre-trained for image classification to generate full-colour channels for the input image which mainly applies in analysing visual imagery. They can be also stated as shift invariant or space invariant artificial neural networks or SIANN, which depends on their shared weights architecture and translation invariance characteristics. Its applications include image and video recognition, recommender systems, medical image analysis, communication processing, image classification, brain-computer interfaces, financial statistics and so on.

Imagenet dataset

The ImageNet project is known as a large visual database constructed to be used for visual object recognition software research. ImageNet is an in-progress research effort to provide researchers with easily accessible image datasets. They can provide more than 14 million images by the project to indicate what objects are pictured and in at least one million of the images, bounding boxes are also provided. ImageNet contains more than 20,000 categories. The datasets from third-party images in the form of URLs are freely provided directly from ImageNet, but the actual images are not owned by ImageNet.

IV. TRAINING THE NETWORK

A VGG-style network with multiple convolutional blocks is what the architecture proposed. There are two or three convolutional layers of every block together with a Rectified long measure and increase to a layer of Batch Normalization.



(FIG2)Convolutional Neural Network

The main goal of the training process is to minimize the loss over the training set. In the colourization problem, the training data contains thousands of colour images and the greyscale versions of them. Here we will be mapping $\hat{Y}$ =F(X) to the two related colour channels Y $\in$ R^{HxWx2}, where H and W are image dimensions, given an input lightness channel X$\in$ R^{HxWx1}. Euclidean loss L2 to compare the predicted and ground truth colours:

$$L_2(\hat{Y}, Y) = 1/2 \sum_{h,w} \left| Y_{h,w} - \hat{Y}_{h,w} \right|_2^2$$

Even if we are tempted to use the above loss function, this process will give really dull image outputs, which will not be preferred.

To solve this problem of dull images we change the loss function. We address the issue by categorizing it as 3 or more groups. We take the output space and break it up into 10 grid size discrete bins. The picture of the input is rescaled to 224x224. This image can be depicted as X. It is converted to $\hat{Z}$ by the neural network as it progresses through the neural network. Mathematically, this network transformation of G can be written as

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

The dimensions of $\hat{Z}$ are HxWxQ, where the height and width of the output of the last convolution layer are H(=56) and W(=56). $\hat{Z}$ contains a vector of Q(=313) values for each of the HxW pixels where each value represents the likelihood of the pixel belonging to that class. For each probability distribution, our goal is to find a single pair of ab channel values $\hat{Z}_{h,w}$. Here we do mapping of Z = G(X) to a probability distribution over possible colours $\hat{Z} \in$ [0, 1] ^(H×W×Q), for a given input X, where Q is the number of ab values quantized.

We will describe the function to evaluate the predicted Z against the ground truth. We need to convert all colour images in the training set to their corresponding Z values. In the training set to their corresponding Z

values, we need to convert all colour images. Mathematically, we just want to reverse the mapping of H.

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

We can simply find the nearest ab bin for every pixel, $Y_{h,w}$ of an output image Y and represent $Z_{h,w}$ as a one-hot vector in which we assign 1 to the nearest ab bin and 0 to all the other 312 bins. But the 5-nearest neighbours are taken into account for a better outcome and a Gaussian distribution is used to measure the distribution $Z_{h,w}$ depending on the distance from the truth of the field.

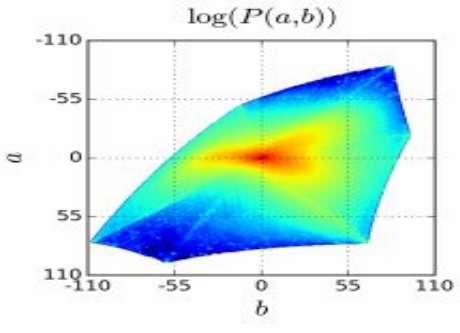
We then use multinomial log loss that calculates our classification model's performance where values between 0 and 1 can be the output. This is done to compare the truth with the estimate.
Lcl(·, ·), characterized as:

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

The above loss function, unfortunately, produces very dull colours. This is because the distribution of colours in ImageNet is heavy around the grey line.

Multinomial loss rebalance

The 1.3 million images were broken down using the Lab model of ImageNet objects and scenes and were used as an input function ('L') and classification marks ("a and "b"). We see that most of the pixels are centred in the centre, where they desaturate the colours. Therefore without taking this into account, the forecasts of the network tend to be desaturated or bland.



(FIG3): demonstrates the empirical distribution of ab-space pixels, obtained from 1.3M ImageNet training images.

By reweighting the loss of each pixel at training time based on the class imbalance issue, we account for the class imbalance problem of lack of pixel colours. This correlates asymptotically to the standard approach resampling of the room for training.
Because of this, we add to our training objective a class re-balancing term and this resamples the rare colours so that they are more represented within the training set than their natural representation. Adding the rebalancing term gives us a more qualitatively and colourful result.

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

Here v is a weighting term that can be used to rebalance the loss based on colour-class rarity, Z to colour values Y are eventually mapped to the probability distribution.

As shown in the *FIG2* we start with a VGG network, then remove the FC layers, then use dilated convolutions to feature extra space within the bottleneck, add additional convolution layers on top and map the features into a predicted distribution for every pixel. The final step is to go from the predicted distribution into a single point estimate. To do this we take an interpolation between the mean and the mode which allows keeping the vibrancy of the output colour while maintaining some spatial consistency.

We define H, which maps the expected distribution of $\widehat{Z}$ to the point estimate $\widehat{Y}$ in the ab space. One alternative is to take the predicted distribution mode for each pixel, this gives a vibrant outcome, but also spatially inconsistent, On the other hand, taking the average of the estimated distribution, showing an artificial mauve tone, shows spatially consistent yet desaturated results. To get the best out of both methods by re-adjusting the T temperature of the softmax distribution and taking the average of the result.

$$H(Z_{h,w}) = E[f_T(Z_{h,w})]$$

$$f_T(z) = (exp(log(z)/T)) / \sum_{q} exp(log(z_q)/T)$$

Setting T = 1 leaves the distribution unchanged, which decreases the temperature. A stronger peak distribution is provided by T, and setting T tends to 0 results in a 1-hot delivery mode encoding. We found the temperature to be T = 0.38.
The CNN is finally trained to map from a grayscale input to a distribution over quantized colour value outputs using this architecture. The ab pair corresponding to the annealed-mean of the $\widehat{Z}_{h,w}$

distribution is interpreted in $\widehat{Y}_{h,w}$, which can be written as a transformation of the original $\widehat{Z}_{h,w}$.

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

Its scale decreases to 56X56 as the picture passes through CNN. Therefore the expected ab image, $\widehat{Y}$, also has a 56X56 dimension. It is sampled to the original image size to retrieve the colour image and then applied to the lightness channel, L, to determine the final colour image.

Lists of the layers used in our architecture during training time are shown below.

Layer	X	C	S	D	Sa	De	BN	L
data	224	3	-	-	-	-	-	-
conv1_1	224	64	1	1	1	1	-	-
conv1_2	112	64	2	1	1	1	✓	-
conv2_1	112	128	1	1	2	2	-	-
conv2_1	56	128	2	1	2	2	✓	-
conv3_1	56	256	1	1	4	4	-	-
conv3_2	56	256	1	1	4	4	-	-
conv3_3	28	256	2	1	4	4	✓	-
conv4_1	28	512	1	1	8	8	-	-
conv4_2	28	512	1	1	8	8	-	-
conv4_3	28	512	1	1	8	8	✓	-
conv5_1	28	512	1	2	8	16	-	-
conv5_2	28	512	1	2	8	16	-	-
conv5_3	28	512	1	2	8	16	✓	-
conv6_1	28	512	1	2	8	16	-	-
conv6_2	28	512	1	2	8	16	-	-
conv6_3	28	512	1	2	8	16	✓	-
conv7_1	28	256	1	1	8	8	-	-
conv7_2	28	256	1	1	8	8	-	-
conv7_3	28	256	1	1	8	8	✓	-
conv8_1	56	128	.5	1	4	4	-	-
conv8_2	56	128	1	1	4	4	-	-
conv8_3	56	128	1	1	4	4	-	✓

(FIG4) X is the space-based resolution of output, C is the number of channels of output; S gives computation stride, values greater than 1 indicate downsampling following convolution, values less than 1 indicate upsampling preceding convolution;D gives the kernel dilation; Sa is the accumulated stride across all preceding layers; De is effective dilation of the layer with respect to the input; BN is if BatchNorm layer was used after layer or not; L is if a 1x1 conv and cross-entropy loss layer were imposed or not.

The efficient dilation of the convolutional kernel is increased through each convolutional block from conv1 to conv5. Efficient dilation is reduced from conv6 through conv8. It is not that straightforward to educate this network as predicted. While after going through the same colouring process, the result may be a realistic picture, it may not be the same as the ground truth. This is because of the occurrence of some type of regression with L2 loss between the estimated and the corresponding labels.

V. METHODOLOGY

The whole process was to take the L channel image and then predict its a and b channels. Combining the prediction and input greyscale would give us the colourized image which can be converted back to the RGB colour space. The algorithm and therefore the approach is already mentioned. Our algorithm uses a Convolutional Neural Network to research the colours across a group of colour images, and their black and white versions. The network is trained on the 1.3M images from the ImageNet training set Any colour image are often changed to grayscale, so paired with its colour version to form a picture for training the model.

We will be using the L channel as the input to the network and train the network to predict the ab channels. We have used LAB colour space. Lab conceals colour information differently. Since the L channel encodes only the vehemence, we will use the L channel as our grayscale input to the network. From there the network is being trained to predict the a and b channels.

Combine the input L channel with the anticipated ab channels. Given the input L channel and therefore the predicted ab channels, we are able to then form our final output image

VI. Result and Conclusion

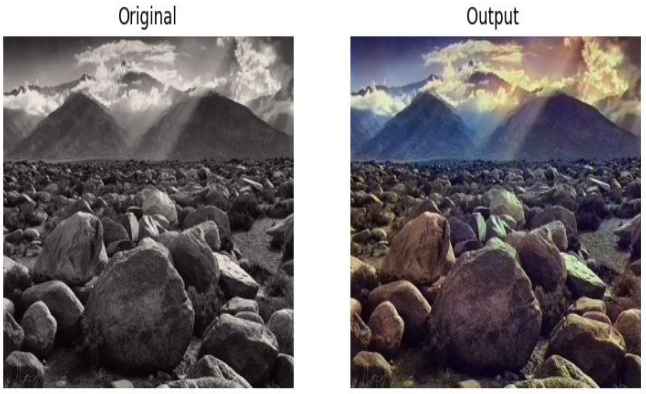
Colour rebalancing renders many pictures very lively and vivid, as we can see. Plausible colours are the bulk of them. On the other hand, it may often add some unnecessary saturated colour patches to some images as well. Bear in mind that there may be several possible solutions as we try to transform a grayscale picture to a colour image. So the way to judge good colourization is not how well it suits the reality of the ground, but how plausible and friendly it appears to the eyes of humans.

Even though the model needs unique types of images to show a perfect image, when the low-level images are provided as input, our model can also produce good colourizations. For the datasets on which it was not directly trained on the model generalizes well. In our model, the aim is more than restoring the true colour of ground reality, creating a convincing colourization that could very easily deceive human observers. It may also lead to human assumptions where the picture is more convincing than the fact of the ground. Our ultimate goal was to build outcomes that are convincing for an individual. Rebalancing colour makes many pictures very bright and colourful. Even then, patches in the images can also be formed by colourization.

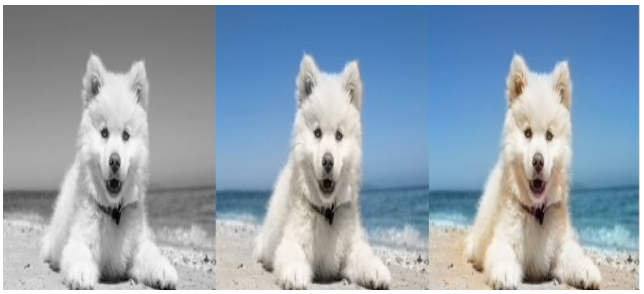
Although image colourization is a challenge for boutique computer graphics, it is also an example in computer vision of a difficult pixel prediction problem.

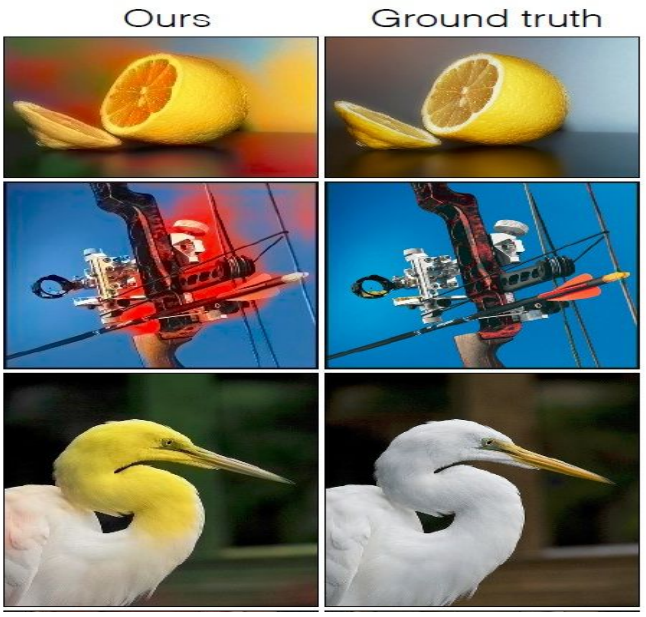
We have done the process of colourization with a deep CNN and a well-chosen objective feature was seen. Indistinguishable from real colour images, results will come closer to producing results. Not only does our method have useful graphics performance, but it can also be displayed as an excuse task for learning representation.
Few images tested by our system are:

‘’



(FIG5)
Examples of coloured image output and the ground truth are:-





As shown in the above images, images like animal and bird groups are often coloured differently from ground reality, as seen in the above pictures, resulting in misclassification of the related species. Notice that while they cause myths, the colourizations are also visually realistic.
Although image colourization may be a task for boutique special effects, it is also an example of a difficult computer vision pixel prediction problem.
We have shown that deep CNN colourization and a well-chosen objective function can be closer to producing real colour photo results.
Not only does our method have a useful graphical output, but it can also be used as a pretext task for learning representation. While only colour-trained, our network learns a representation that is remarkably useful compared to other self-supervised pre-training methods for object classification, detection, and segmentation, performing strongly.

The model also does a very good job of portraying the blue sky and green foliage in outdoor scenes. The model also forecasts an orange sky, given the outline of a tree, suggesting the fact that it has captured the notion of sunset.

REFERENCES

[1] Cheng, Z., Yang, Q., Sheng, B.: Deep colorization. In: Proceedings of the IEEE

[2] Charpiat, G., Hofmann, M., Schölkopf, B.: Automatic image colorization via multimodal predictions. In: Computer Vision–ECCV 2008.

[3] Cheng, Zezhou, Qingxiong Yang, and Bin Sheng. "Deep colorization." Proceedings of the IEEE International Conference on Computer Vision. 2015.

[4] Huang, Yi-Chin, et al. "An adaptive edge detection based colorization algorithm and its applications." Proceedings of the 13th annual ACM international conference on Multimedia.ACM, 2005.

[5] Deep Learning based image colorization with OpenCV

https://cv-tricks.com/opencv/deep-learning-image-colorization/

[6] Zhu, Jun-Yan, et al. "Unpaired image-to-image translation using cycle-consistent adversarial networks." arXiv preprint arXiv:1703.10593 (2017).

[7] Dahl, Ryan. "Automatic colorization." (2016)

[8] K. Hornik, M. Stinchcombe, and H. White. Multilayer feedforward networks are universal approximators. Neural networks, 2(5):359–366, 1989. 3

[9] Y.-C. Huang, Y.-S. Tung, J.-C. Chen, S.-W. Wang, and J.-L. Wu. An adaptive edge detection based colorization algorithm and its applications. In Proceedings of the 13th Annual ACM International Conference on Multimedia, MULTIMEDIA ’05, pages 351– 354, 2005. 1, 2

[10] R. Irony, D. Cohen-Or, and D. Lischinski. Colorization by example. In Eurographics Symp. on Rendering, volume 2. Citeseer, 2005. 1, 2, 5, 6, 7

[11] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In Advances in neural information processing systems

[12] A. ertzmann, C. E. Jacobs, N. Oliver, B. Curless,

and D. H. Salesin. Image analogies. In Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH ’01, pages 327– 340, 2001. 2 H