A learned representation of artist-specific colourisation

Nanne van Noord Eric Postma
Cognitive Science and Artificial Intelligence group, Tilburg University
Warandelaan 2, Tilburg, Netherlands
{n.j.e.vannoord, e.o.postma}@tilburguniversity.edu

Abstract

The colours used in a painting are determined by artists and the pigments at their disposal. Therefore, knowing who made the painting should help in determining which colours to hallucinate when given a colourless version of the painting. The main aim of this paper is to determine if we can create a colourisation model for paintings which generates artist-specific colourisations. Building on earlier work on natural-image colourisation, we propose a model capable of producing colourisations of paintings by incorporating a conditional normalisation scheme, i.e., conditional instance normalisation. The results indicate that a conditional normalisation scheme is beneficial to the performance. In addition, we compare the colourisations of our model that is trained on a large dataset of paintings, with those of competitive models trained on natural images and find that the painting-specific training is beneficial to the colourisation performance. Finally, we demonstrate the results of stylistic colour transfer experiments in which artist-specific colourisations are applied to the artworks of other artists. We conclude that painting colourisation is feasible and benefits from being trained on a dataset of paintings and from applying a conditional normalisation scheme.

1. Introduction

Image colourisation is the task of hallucinating a colour image given a greyscale image. This task is clearly underconstrained in that a pixel with a given greyscale value can be assigned a number of different colours. Nonetheless, for most natural images there are colours which are much more likely than others, e.g., given a tropical beach scene we can all imagine that the sky and water are blue, the sand a light tan, and the palm leaves green. In other words, the semantics of the image region impose constraints on what would be plausible colours. If we are able to recognise *what* is depicted, we may be able to suggest a plausible colourisation. Recent work has shown that Convolutional Neural Networks (CNN) can obtain sufficient visual understanding to perform automatic image colourisation [15, 28, 3, 8, 10].

(a) "*View in the Woods*" (b) "*Evening; Red Tree*"

Figure 1. Examples of two paintings depicting a similar scene, but with very different colour usage. Left is "*View in the Woods*" ("*Bosgezicht*") by Jan van Kessel (courtesy of the Rijksmuseum) and right "*Evening; Red Tree*" ("*Rode boom*") by Piet Mondrian (courtesy of the Gemeentemuseum Den Haag).

Depending on the type of image other factors than the image semantics might play a role in determining the likelihood of colours. For paintings the idiosyncratic use of colours by the artist greatly influences the likelihood of colours. While (realistic) paintings are often intended as realistic representations of natural scenes, the geographical, historical, and economical availability of colourants might have restricted the artist's use of colour. Additionally, and maybe more important to painters; their choice of colours is often guided by aesthetic considerations [16]. As such we pose that due to the inherent complexity of colouring paintings it is necessary to take into account both the image semantics, and the artist's palette. An example of the influence the artist's palette has on the used colours can be seen in Figure 1, showing two similar scenes, one with realistic colours and the other with seemingly unrealistic colours.

An image colourisation model might learn to take the artist's palette into account in the following two ways. The first way of taking the artist's palette into account is by acquiring a model of the artist's style. Previous work has shown that CNNs are capable of acquiring a model of the artist's style [26]. Therefore, the model could learn to recognise which visual content is artist-specific, and use this to facilitate artist-specific colourisation. The second way of

taking the artist's palette into account, is to condition (part of) the CNN on the artist, and explicitly enforce that it acquires an artist-specific mapping.

In this paper, we compare these two approaches for producing artist-specific colourisations of paintings. Our results indicate that explicitly conditioning the network makes it possible to influence the colourisation, but that surprisingly even without this explicit signal the network is able to hallucinate plausible colours.

The remainder of this paper is organised as follows. Section 2 reviews previous work on image colourisation, normalisation, and computational art analysis. In Section 3 we describe the details of our approach. Followed by Section 4 in which the results are presented, as well as a number of qualitative comparisons of the colourisation results for various models. In Section 5 we discuss several questions which arose during this work. Finally in Section 6 we conclude by stating that the approach presented is capable of producing highly diverse visually appealing colourisations of paintings.

2. Previous work

This section reviews earlier work pertaining to our colourisation approach: image colourisation, normalisation, and computational art analysis.

2.1. Image Colourisation

Work on image colourisation can be divided into user-based approaches and fully automatic approaches. User-based approaches rely on interaction (e.g., provide *scribbles* or reference images) with the user, whereas fully automatic approaches aim to provide a colourised image without user interaction, see [1] for a comprehensive overview.

Recent work on fully automatic image colourisation has shown that Convolutional Neural Networks (CNN) are capable of producing visually appealing colourisation results [15, 28, 3, 8, 10]. CNN-based fully automatic approaches can be categorised into two groups: (1) per-pixel descriptor approaches [2, 15] and (2) encoder-decoder type architectures [8, 28, 3, 10]. The per-pixel descriptor approach consists of passing the input image through a (pretrained) CNN and extracting a hypercolumn descriptor [7] for each pixel. The per-pixel descriptors are subsequently fed to a classifier that predicts the colour based on the descriptor. Hypercolumns describe the region around the pixel at different scales, incorporating a large amount of context, which results in accurate predictions. However, densely extracting hypercolumns from an image is very memory intensive, making it expensive to train an end-to-end system. Larsson et al. [15] propose to extract the hypercolumns from a subset of randomly chosen locations, but only show that this works for fine-tuning a network, not for training a network from scratch.

In contrast, so called encoder-decoder architectures have shown very promising results when trained from scratch [10]. Typically, this type of architecture consists of an encoder which follows a traditional CNN layout, i.e., several layers which have an increasing number of filters and a decreasing spatial resolution. Followed by a decoder which either upsamples using interpolation (e.g., nearest-neighbour, bilinear, or bicubic), or *deconvolution* (i.e., fractional strided convolution) [27]. Encoder-decoder architectures are trained in either a Generative Adversial setting [10], or with a pixel-wise loss [8, 28, 3].

2.2. Normalisation

Most modern CNN make use of Batch Normalisation (BN) for each nonlinear unit in the network. BN reduces *internal covariate shift* (changes in the distribution of the inputs for a layer, due to weight updates in preceding layers) and accelerates training [9]. Given a batch of size T, BN normalises each channel c of its input $x \in R^{T\times C\times W\times H}$ such that it has zero-mean and unit-variance. Formally, BN is defined as:

$$y_{tijk} = \gamma_i \left(\frac{x_{tijk} - \mu_i}{\sigma_i} \right) + \beta_i. \tag{1}$$

where μ_i and σ_i describe the mean and standard deviation for channel C_i across the spatial axes W and H, and the batch of size T Additionally, for each channel there is a pair of learned parameters γ and β, that scale and shift the normalised value such that they may potentially recover the original activations if needed [9]. BN is applied in a different way training and testing. Ideally we would calculate μ_i and σ_i on the whole dataset prior to training, but as they depend on the incrementally learned weight values of preceding layers this is not possible. Instead, during training μ_i and σ_i are calculated on the actual batch and added to moving averages. The resulting averages are used during testing.

In recent work on style transfer, it was shown that accounting for instance-specific contrast improves generation results [25]. The approach, called Instance Normalisation (IN), modifies BN in the following two ways: (1) IN calculates μ_i and σ_i for each specific instance rather than for the entire batch as in BN. (2) IN does not maintain moving averages, and is applied identically during training and testing. We expect that IN might also be beneficial for painting colourisation, or even image colourisation in general, because uniform contrast changes should not alter the colourisation substantially. Moreover, a dataset of paintings consists of samples generated from different distributions (i.e., painters), as such we expect it is very unlikely that a single mean and variance are sufficient to adequately normalise the activations without introducing artifacts.

More recently, there has been work on extending feed-forward style transfer [12] to deal with multiple styles by conditioning the shifting and scaling parameters on the style [4]. Conditional Instance Normalisation (CIN) modify IN such that the γ and β parameters are $N \times C$ matrices rather than length C vectors, where N is equal to the number of styles being modelled. In this work we will use CIN to modify the colour use of different artists, by conditioning the shifting and scaling parameters on the artist.

2.3. Computational art analysis

There is large body of work on the computational analysis of artworks, while a large portion of this work is concerned with learning characteristics of artists for classification [11, 13, 26], an increasing body of work is emerging which tries to capture artist-specific characteristics for generative purposes [5, 25, 4]. This latter type of work, is generally concerned with *style transfer* (i.e., given a style image S and a content image C produce a single image with style S_{style} and content $C_{content}$). In this work we are only concerned with the colour aspects of the style.

2.4. Our Contributions

In this work we make the following three contributions: (1) We present an image colourisation model[1] building on components from previous works, which we apply and evaluate on a dataset of paintings. (2) We compare various normalisation schemes, investigating the influence of batch versus instance normalisation, and conditional versus unconditional normalisation. (3) We show that the models using conditional and 'unconditional' instance normalisation utilise their visual understanding of image regions in an artist-specific way, resulting in visually appealing and diverse colourisations of paintings.

3. Method

In this work we use a *'encoder-decoder'*-style convolutional neural network to perform end-to-end colourisation of paintings, with the additional goal of learning the artist's unique palette. To explicitly learn the artist's palette, or colour use, we add Conditional Instance Normalisation (CIN) to the network, where the γ and β parameters are conditioned on the artist.

3.1. Loss

For image colourisation the goal is to learn a mapping $\hat{Y} = F(X)$ from a greyscale image $X \in \mathbb{R}^{H \times W}$ to a colour image Y, where the pixel lightness values are taken to represent the greyscale image, and H, W are the image width and height respectively. Typically colour images are represented in RGB colour space that combines colour information with luminance (intensity) information, luminance is encoded in the mean of the R, G, and B channels.

For image colourisation the CIE Lab colour space is more appropriate, because it represents luminance (L) as a channel separate from the two colour channels **a** and **b**. Colourisation in **Lab** colour space means mapping the **L** channel of an image to the **Lab** channels. In CIE Lab, **a** represents colours along the red-green axis and **b** along the blue-yellow axis. Both CIE Lab colour values are continuous valued. Hence, colourisation could be formulated as a regression task. However, previous work has shown that formulating colourisation as a regression task tends to result in desaturated colours [15, 28]. This is most likely due to the tendency of regression to favour the mean when dealing with a multimodal distribution across colours, i.e. if a colour regression model is trained on a database of t-shirts, where half of the t-shirts are completely white, and the other half are completely black it will probably favour grey at test time.

A common solution to deal with this limitation of regression is to reformulate the task as a classification task, by discretising the target, and effectively predicting a histogram across colour bins for each pixel. We discretise the **a** and **b** channels separately by binning the axes with Q equal-width bins, where we set $Q = 32$ following [15]. Therefore, Y becomes a four dimensional matrix $Y \in [0, 1]^{H \times W \times Q \times 2}$, and the loss effectively becomes the sum of the cross entropy loss for both the **a** and the **b** channel.

3.2. Class rebalancing

Zhang et al. [28] show that during training it is possible to re-weight the loss at each pixel, following an approach akin to sample weighting. The loss at each pixel is re-weighted based on a weighting factor determined by the rarity of the target colour. This approach prevents the loss function from being dominated by highly common colours and is similar to the approach described in [3].

Following the procedure describe in [28] we estimate the empirical probability distribution of colours in the discretised space $p \in \Delta^Q$ on the training set, which is smoothed with a Gaussian kernel G_σ. Subsequently, the contribution of the probability-weighted distribution is parameterised by $\lambda \in [0, 1]$. More formally, Zhang et al. [28] define the weighting factor $w \in \mathbb{R}^Q$ as:

$$w \propto \big((1 - \lambda)(G_\sigma \circ p) + \lambda\big)^{-1} \tag{2}$$

Unlike [28] we have discretised the **a** and **b** channels separately, therefore we also have separate losses for the **a** and **b** channels. Subsequently, we weight the channels independently using weighting factors w_A and w_B respectively. We used the values of $\lambda = \frac{1}{2}$ and $\sigma = 5$ following [28].

[1] https://github.com/Nanne/conditional-colour

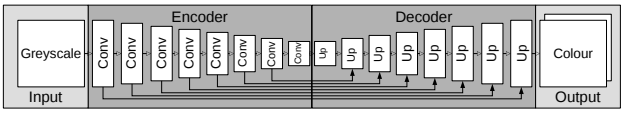

Figure 2. Visualisation of the network architecture. *Conv* refers to a convolution layer, and *Up* to an upsampling layer. The network input is 224×224 and the output is $224 \times 224 \times 2Q$. The bottom arrows between matching layers in the encoder and decoder indicate skip connections. Skip connections differ from regular connections in that they are concatenated to the output of the matching layer, integrating lower-level features at a higher spatial resolution to upsampled higher-level features.

3.3. Network architecture

The network architecture used for our colourisation model is based on the "U-Net" architecture[19] used in [10], and is shown in Figure 2. The U-Net architecture is an encoder-decoder architecture with skip connections between matching layers in the encoder and the decoder. The skip connections enable a direct mapping between layers at the same spatial scale. This allows the encoder-decoder path of the network to model the mapping from the grey values to colours, without being responsible for a reconstruction of all image details. We modified U-Net by replacing the upsampling (de)convolution layers with upsampling by means of nearest-neighbour interpolation, followed by a convolutional layer, as described in [4]. This upsampling method helps to avoid high spatial frequency noise [4] and 'Checkerboard' artifacts [17]. The kernel size for all convolutional layers was set to 4×4, and all convolutional layers in the encoder use a stride of 2. All layers use a ReLU nonlinearity, except the last layer which is followed by a softmax activation function.

The network outputs a colour histogram for each pixel, to convert this to an actual colour we take the 'expectation' over the histogram i.e., the weighted sum of the colour bins [15]. This results in smooth colour transitions and avoids the discontinuities obtained when taking the colour of the highest bin.

3.4. Training details

For training we use ADAM [14] ($\alpha = 0.001, \beta_1 = 0.9, \beta_2 = 0.999$), and all the weights are initialised using Xavier weight initialisation [6]. In terms of data augmentation we perform random horizontal flips, take 224×224 pixel crops, and introduce a random uniform brightness shift on the **L** channel in the interval $[-d, d)$. The value of d was chosen to be smaller than noticeable to human observers i.e., the colour difference (ΔE) was smaller than 1 [23].

4. Experiment

To evaluate our colourisation model we compare the performances of the following seven approaches on a painting dataset:

1. **Greyscale** - Baseline using greyscale versions of images (i.e., original L channel and ab channels set to 0).
2. **Larsson et al. [15]** - A CNN based approach using sparse hypercolumns trained on natural images.
3. **Zhang et al. [28]** - An encoder-decoder style network trained on natural images and paintings.
4. **Ours BN** - Our model using Batch Normalisation trained on paintings.
5. **Ours IN** - Our model using Instance Normalisation trained on paintings.
6. **Ours CIN** - Our model using Conditional Instance Normalisation trained on paintings, conditioned on 1.678 artists.
7. **Ours randomised-CIN** - Our model with Conditional Instance Normalisation, using a random artist rather than the actual. If conditioning on the artist works then we would expect this to perform worse than our CIN model.

For each of the seven approaches, we compute the micro-averaged root mean square error (RMSE) across all pixels in **ab** space, and macro-averaged the peak signal to noise ratio (PSNR) in RGB space per image. The greyscale approach functions as a baseline by providing no colourisation, i.e. all zero **ab** values.

The second and third approach (by Larsson et al. and Zhang et al.), are originally trained on a dataset of natural image (the ImageNet dataset) [20], and not on paintings. Both approaches incorporate copies of the first layers from a trained VGG-16 model [24], and are state-of-the-art (natural) image colourisation models. To compare the influence of the training data we fine-tune model[2] by Zhang et al. [28] on our painting dataset. There are two motivations for fine-tuning, (1) the performance of the models trained on natural images show how well such models generalise to paintings. (2) Fine-tuning the model allows us to compare the benefits of training on paintings, and how our model compares to this model in a comparable setting. For the four variations of our model the scores reveal the effectiveness of the different normalisation schemes, where the randomised-CIN is used as an extra validation of the CIN model. If the randomised-CIN model performs worse than the CIN model

[2]We were unable to perform any type of training with the model by Larsson et al.

we can infer that the conditioning is effective. In addition, we perform qualitative evaluations of the best performing colourisation approach and demonstrate the transfer of the colour style of one artist onto an artwork of another artist.

In the remainder of this section we will introduce the dataset used for the experiment, and present the results the different approaches obtain.

4.1. Painting colourisation dataset

The painting colourisation performances is evaluated on the "*Painters by Numbers*" dataset as published on Kaggle[3]. This dataset is a collection of images collected from different sources, though the majority was retrieved from "*Wikiart*" a repository which was used in a number of previous publications involving computational artwork analysis [21, 22].

A portion of the images included in this dataset are colourless or contain very little colour. For most of these images this is because they are drawings on paper, and while the paper might not be purely white, a greyscale prediction would often be very close to the ground truth. Nevertheless, we chose to keep these images in the dataset as we feel they are inherent to the task, and fine-tuning the cut-off point for how much colour is desirable might arbitrarily influence the task.

From the "*Painters by Numbers*" dataset we select the subset of artists who have at least 5 artworks in the dataset, which results in a dataset consisting of 101.580 photographic reproductions of artworks produced by a total of 1.678 artists. Subsequently we divide the dataset into a training, validation, and test set used for training the model, evaluating stopping criteria, and reporting evaluation performances respectively. Both the test and validation set consist of 5000 images obtained by stratified random sampling.

4.2. Painting colourisation

In this section the results on the main image colourisation task in this work are described. All results are measured using the micro-averaged root mean square error (RMSE) across all pixels in **ab** space, and the macro-averaged peak signal-to-noise ratio (PSNR) across images in RGB space.

Results of the comparison between the seven approaches described in Section 4 in Table 1 show that our model achieves the highest performance according to RMSE. On this dataset all models score below the PSNR baseline, despite our model achieving the highest performance of all models. We suspect that the high PSNR for the baseline is an artifact of the colourless images in the dataset, and the calculation of this metric in RGB space. Nevertheless, we pose that the metric remains useful to compare performance between approaches.

[3] https://www.kaggle.com/c/painter-by-numbers

Table 1. Painting colourisation results measured using RMSE across all pixels, and PSNR in RGB space. The goal is to have a low RMSE, and a high PSNR. "Greyscale" is a baseline which provides no colourisation.

Method	RMSE	PSNR
Greyscale	0.175	**24.66**
Trained on natural images		
Larsson et al. [15]	0.168	22.18
Zhang et al. [28]	0.163	22.29
Fine-tuned on paintings		
Zhang et al. [28]	0.175	21.65
Trained on paintings		
Ours BN	0.146	23.26
Ours IN	0.149	23.31
Ours CIN	**0.145**	**23.34**
Ours Randomised CIN	0.164	22.31

Our model outperforms the baseline regardless of the normalisation scheme, and it outperforms the two previous colourisation approaches (by Larsson et al. and Zhang et al.) regardless of whether they were trained on natural images or fine-tuned on paintings. Nonetheless, there are differences in performance between the normalisation schemes. With CIN performing slightly better than IN and BN, both in terms of RMSE and PSNR. Moreover, from the comparison between CIN and randomised-CIN we can learn that conditioning on the correct artist is important, in that using a random artist results in a deteriorated performance, which demonstrates that the CIN model learns to colourise in an artist-specific manner.

For a qualitative comparison between our models we show three sets of the colourisation results, the first set in Figure 3 shows the best case performance, the second set in Figure 4 the worst case, and the third set in Figure 5 the expected performance. These sets were created based on the RMSE obtained by the best performing model (Ours CIN). In Figure 3 we show the colour paintings in the best case. The best performances were obtained for a few natively greyscale paintings/drawings contained in the dataset. These will be discussed separately. The presence of these paintings/drawings is presumably also the cause for the high PSNR for the greyscale baseline.

When comparing the colourisations in Figure 3 we can observe that all three normalisation schemes produce plausible colourisations, despite not always exactly matching the ground truth. It appears that the IN and BN model produce colours which are more typical for the entire dataset, whereas CIN produces colours which closer match the original: a more saturated red in the first row, greys/silvers instead of browns in the third row, and a yellow sky rather than a blue sky in the last row. These results are in line with what we would expect as differences between these models.

The cases for which we obtain the worst RMSE are those

Figure 3. Example colourisation results on Painters by Numbers. Colour images with lowest RMSE according to our CIN model.

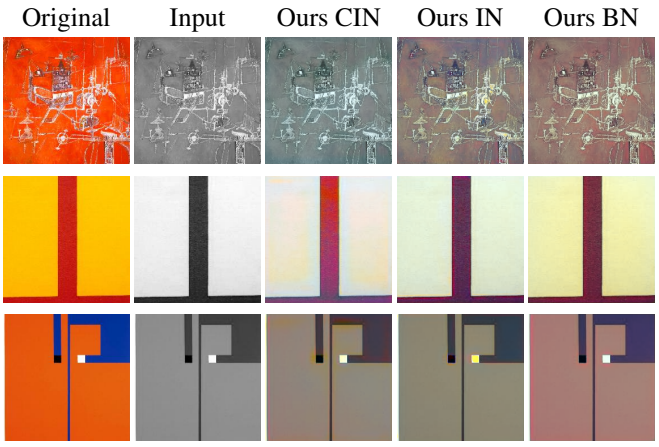

Figure 4. Example colourisation results on Painters by Numbers. Shown examples have the highest RMSE according to our CIN model.

shown in Figure 4. For these (abstract) artworks there appears to be little to no visual semantics that provide clues about the colours used. The experimental use of colour by abstract artists such as Mark Rothko (in the second row) makes colourisation virtually impossible.

In order to see the expected performance of the CIN model we present the images shown in Figure 5, which were randomly sampled from around the median RMSE. These images show that the colourisations for both CIN and IN are very consistent with the original, although all models predict the jacket in the artwork on the second row to be red rather than blue. However, given that there is no indication in the input which colour it should be, and either colour is equally plausible we would consider this a good colourisation. The colourisations produced by BN are not far behind, though they seem to be less spatially consistent.

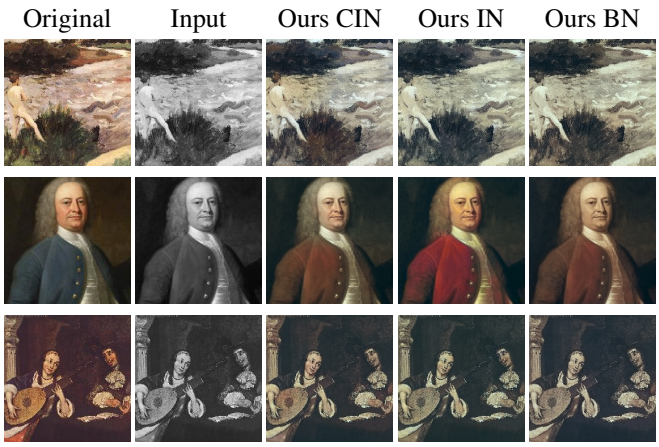

Figure 5. Example colourisation results on Painters by Numbers. Shown examples were randomly sampled from around the median RMSE for our CIN model.

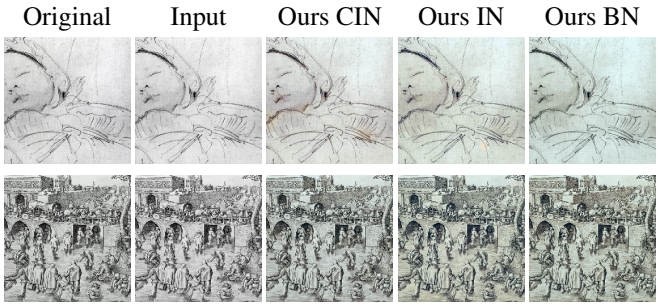

Figure 6. Example colourisation results on Painters by Numbers. Images with lowest RMSE according to our CIN model.

In Figure 3 we showed the colour images for which the CIN model obtained the lowest RMSE. As stated, the lowest RMSE scores were obtained for the natively greyscale images shown in Figure 6. The best hallucination for natively greyscale paintings and drawings, is reproduction of the input input (with potentially a slight uniform hue change). It appears all models are able to learn to generate a greyscale reproduction, though with slight hue differences. In hindsight, we could have removed the natively colourless or almost colourless artworks from the Painters by Numbers dataset to make the colourisation task more consistent.

For qualitative comparison between our best performing model (CIN) and the models by Larsson et al. [15] and Zhang et al. [28] we show three images in Figure 7 for which the absolute difference in RMSE between our CIN model and the Larsson et al. [15] model is the largest. From these images we can observe that this mainly concerns abstract artworks for which a human observer would have difficulty picking the most plausible colourisation. Fortunately, our CIN model has artist-specific information, therefore it can produce a reasonable colour, despite the lack of

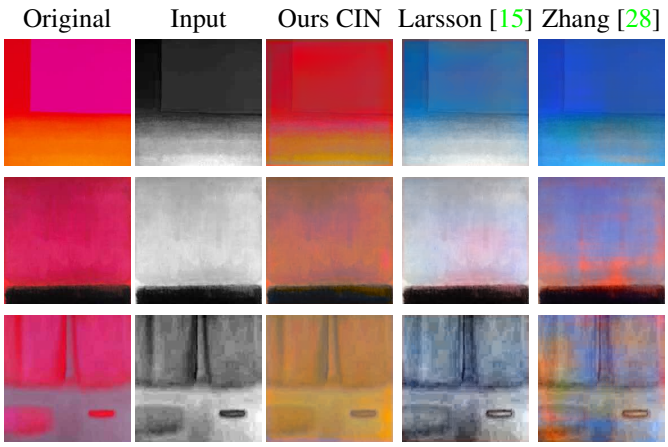

Figure 7. Example colourisation results on Painters by Numbers. Images where our CIN model outperforms [15] with the biggest RMSE difference.

semantic information in the image.

4.3. Stylistic colour transfer

In the previous section we have shown that the performances of normalisation schemes are very similar. For generative purposes, the CIN model has an additional advantage in that we can choose in which colour style to render the artwork. As a result, we can transfer the colour style of one artist onto an artwork of another artist. In this section we perform a qualitative comparison of a number of artworks on which we applied stylistic colour transfer. As the sources for our colour transfer experiments, we selected the colour styles of Maria Primachenko and Mark Rothko, because of their prominent use of colour. Note that this approach differs from what is commonly referred to as colour transfer, in that we learn the style of an artist from a database of images, rather than from a single reference image [18].

The stylistic colour transfer visualisations can be found in Figure 8. These columns (from left to right) show the greyscale input to the model, the original artwork in colour, a colourisation produced conditioned on the actual artist, a colourisation conditioned on Maria Primachenko, and a colourisation conditioned on Mark Rothko.

The first row shows an artwork by Roy Lichtenstein. The colourisation conditioned on his colour style is not very close to the original. Still, it does match the colour palette of many of his other artworks. The colourisation conditioned on Maria Primachenko is much more yellow, with some purple highlights. The colourisation conditioned on Mark Rothko is mainly in shades of red and orange. A similar pattern can be observed in the next rows, for the colourisations of an artwork by Marc Chagall, and one artwork by Louisa Matthiasdottir.

For all artworks we can observe that the three colourisations differ strongly, illustrating the artist-specific effect of the CIN model.

5. Discussion

The main aim of this paper was to determine if we can create a colourisation model for paintings which can deal with the inherent complexity of the task due to the influence of both image semantics and the artist's palette. Our results indicate that automatic colourisation models can produce plausible colourisations for paintings, and that performing the colourisation in an artist-specific manner appears beneficial. In what follows, we discuss (1) artist-specific colourisation, (2) normalisation schemes, (3) the use of paintings (rather than natural images) for training a colourisation model, and (4) evaluation of painting colourisation models.

(1) Artist-specific colourisation. We aimed to learn a representation of the artists colour usage such that we could do artist-specific colourisation. We compared an approach to do this explicitly (CIN) with two approaches which might be able to do this implicitly (BN and IN). Our results show that while the CIN approach can be used to explicitly alter the colourisation, the IN (and to a lesser extent the BN) approach appear to recognise the artist and use this as an information source for the colourisation. Therefore, we pose that the minor difference in performance between CIN and IN is due to the ability of the IN approach to recognise the artist or the art style to a sufficient extent, such that it is not necessary to explicitly pass this as a signal to the network.

(2) Normalisation Schemes. We found the difference in performance between the normalisation schemes to be very small. CIN offers some additional functionality in that we can influence the colourisation, at the cost of extra (conditional) parameters. Moreover, while in the work of [4] CIN is used to achieve impressive style transfer results, we pose that the representational power of the scale and shift parameters in CIN is insufficient to capture the full complexity of an artist's palette. Therefore, the main difference between the normalisation schemes seem to come down to saturation levels and small colour variations. Still, the benefits of CIN are very clear and give a definite improvement in performance for painting colourisation. It would be worthwhile to investigate whether this is the case as well for other image colourisation tasks. This is left to future work.

(3) Use of paintings for training. It could be argued that a painting specific colourisation model is not necessary, as applying realistic colours learned from natural scenes should be sufficient to produce satisfactory results. Our results indicate that the visual structure in paintings is different to such a large extent that image colourisation models trained on natural scenes only generalise to paintings which are (hyper)realistic, and do not recognise the structure in more abstract paintings. Our results indicate that fine-tuning such a network does not help to overcome this,

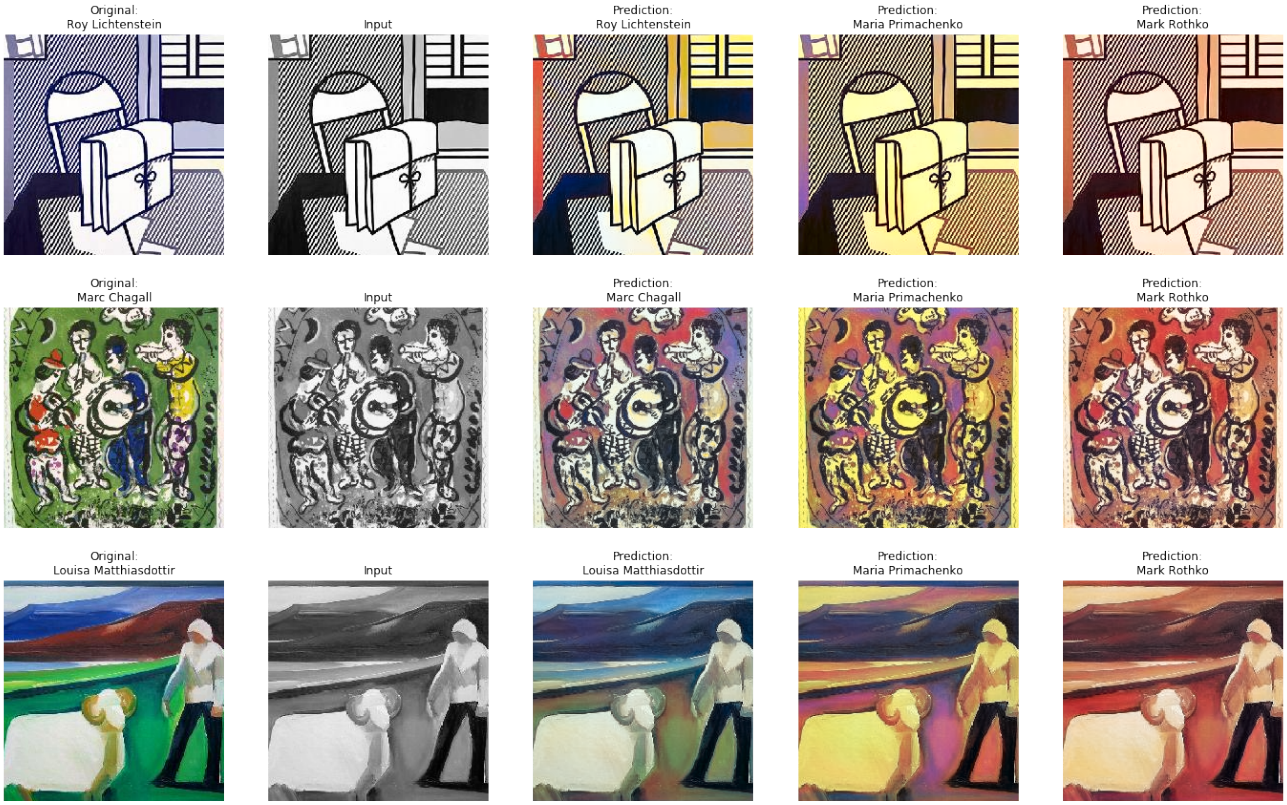

Figure 8. Stylistic colour transfer results. For three greyscale images we show the colourisation results of conditioning on the actual artist (third column) on Maria Primachenko (fourth column), and on Mark Rothko (last column).

rather that it appears to worsen the results. Additionally, besides differences in image structure for abstract paintings, these paintings also tend to use a different palette than found in nature, making it necessary to train a model specifically for this task. Although the model itself could be a generically applicable model, such as the model presented in the current paper.

(4) Evaluation of painting colourisation. A notable problem for image colourisation is how to do the evaluation. While quantitative measures, such as the ones used in this work, given an indication of the performance of the model, they have a number of pitfalls. These pitfalls mainly concern the bias of these measures to prefer greyscale over a wrong colour, even when the saturation levels match the ground truth (i.e., greyscale is preferred over blue when the ground truth is green). To overcome this, a number of works have employed user studies [28, 10], or external evaluation by means of a classification task [28]. For painting colourisation the former is hindered by the presence of abstract paintings for which naive users have difficulty judging the plausibility. The latter approach leads to incomparable results when applied to our work as our conditional model receives information about who the artist is, which might give it an unfair advantage. How to accurately evaluate colourisation models remains an open question.

6. Conclusion

In this work we proposed an image colourisation model capable of producing colourisations of paintings specific to the colour style of an artist. While the model's performance was demonstrated on paintings and artists, we pose that it is a general approach which could be applied to a wide variety of image colourisation tasks, as none of the components are specific to the painting domain. However, we pose that for cultural heritage applications the conditional aspect is most useful, as there is often a creative human component which determines the image appearance. In conclusion, our model is capable of producing plausible colourisations of paintings, and is highly diverse when varying the artist on which the colourisation is conditioned.

7. Acknowledgements

The authors thank the anonymous reviewers for their insightful and constructive comments. The research reported in this paper is part of the REVIGO project, supported by the Netherlands Organisation for scientific research (NWO; grant 323.54.004) in the Science4Arts research program.

References

[1] G. Charpiat, I. Bezrukov, Y. Altun, and B. Hofmann, Matthias Schölkopf. Machine Learning Methods for Automatic Image Colorization. In R. L. Editor, editor, *Computational Photography: Methods and Applications*, pages 1–27. CRC Press, 2011. 2

[2] R. Dahl. Automatic colorization. http://tinyclouds.org/colorize/:, 2016. 2

[3] A. Deshpande, J. Lu, M.-c. Yeh, and D. Forsyth. Learning Diverse Image Colorization. *arXiv preprint*, 2016. 1, 2, 3

[4] V. Dumoulin, J. Shlens, M. Kudlur, G. Brain, and M. View. A learned representation for artistic style. In *arXiv preprint*, 2016. 3, 4, 7

[5] L. A. Gatys, A. S. Ecker, and M. Bethge. A Neural Algorithm of Artistic Style. *arXiv*, pages 3–7, 2015. 3

[6] X. Glorot and Y. Bengio. Understanding the difficulty of training deep feedforward neural networks. *Proceedings of the 13th International Conference on Artificial Intelligence and Statistics (AISTATS)*, 9:249–256, 2010. 4

[7] B. Hariharan, P. Arbeláez, R. Girshick, and J. Malik. Hypercolumns for Object Segmentation and Fine-grained Localization. *Computer Vision and Pattern Recognition (CVPR)*, nov 2015. 2

[8] S. Iizuka, E. Simo-serra, and H. Ishikawa. Let there be Color!: Joint End-to-end Learning of Global and Local Image Priors for Automatic Image Colorization with Simultaneous Classification. *ACM Transactions on Graphics*, 35(4):1–11, 2016. 1, 2

[9] S. Ioffe and C. Szegedy. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. In *ICML*, pages 448–456. JMLR, 2015. 2

[10] P. Isola, J.-y. Zhu, T. Zhou, and A. A. Efros. Image-to-Image Translation with Conditional Adversarial Networks. *arXiv preprint*, 2016. 1, 2, 4, 8

[11] C. R. Johnson, E. Hendriks, I. J. Berezhnoy, E. Brevdo, S. M. Hughes, I. Daubechies, J. Li, E. Postma, and J. Z. Wang. Image processing for artist identification. *IEEE Signal Processing Magazine*, 25(4):37 – 48, 2008. 3

[12] J. Johnson, A. Alahi, and L. Fei-fei. Perceptual Losses for Real-Time Style Transfer and Super-Resolution. In *European Conference on Computer Vision*, 2016. 3

[13] S. Karayev, M. Trentacoste, H. Han, A. Agarwala, T. Darrell, A. Hertzmann, and H. Winnemoeller. Recognizing Image Style. In *British Machine Vision Conference (BMVC)*, nov 2014. 3

[14] D. P. Kingma and J. L. Ba. Adam: a Method for Stochastic Optimization. In *International Conference on Learning Representations*, pages 1–13, 2015. 4

[15] G. Larsson, M. Maire, and G. Shakhnarovich. Learning Representations for Automatic Colorization. In *European Conference on Computer Vision (ECCV)*, 2016. 1, 2, 3, 4, 5, 6, 7

[16] S. M. Nascimento, J. M. Linhares, C. Montagner, C. A. João, K. Amano, C. Alfaro, and A. Bailão. The colors of paintings and viewers' preferences. *Vision Research*, 130:76–84, 2017. 1

[17] A. Odena, V. Dumoulin, and C. Olah. Deconvolution and Checkerboard Artifacts. http://distill.pub/2016/deconv-checkerboard/, 2016. 4

[18] E. Reinhard, M. Ashikhmin, B. Gooch, and P. Shirley. Color Transfer between Images. *IEEE CG&A special issue on Applied Perception*, 21:34–41, 2001. 7

[19] O. Ronneberger, P. Fischer, and T. Brox. U-Net: Convolutional Networks for Biomedical Image Segmentation. *MICCAI*, pages 234–241, 2015. 4

[20] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei. ImageNet Large Scale Visual Recognition Challenge. *Arxiv*, page 37, 2014. 4

[21] B. Saleh and A. Elgammal. Large-scale Classification of Fine-Art Paintings: Learning The Right Metric on The Right Feature. *arXiv*, page 21, 2015. 5

[22] B. Seguin, C. Striolo, and F. Kaplan. Visual Link Retrieval in a Database of Paintings. In *European Conference on Computer Vision (ECCV)*, pages 753–767, 2016. 5

[23] G. Sharma, W. Wu, and E. N. Dalal. The CIEDE2000 Color-Difference Formula: Implementation Notes, Supplementary Test Data, and Mathematical Observations. *Color Research & Application*, 30(1):21–30, 2005. 4

[24] K. Simonyan and A. Zisserman. Very Deep Convolutional Networks for Large-Scale Image Recognition. *ArXiv*, pages 1–14, sep 2015. 4

[25] D. Ulyanov, A. Vedaldi, and V. Lempitsky. Instance Normalization: The Missing Ingredient for Fast Stylization. (2016). 2, 3

[26] N. van Noord, E. Hendriks, and E. Postma. Toward Discovery of the Artist's Style: Learning to recognize artists by their artworks. *IEEE Signal Processing Magazine*, 32(4):46–54, 2015. 1, 3

[27] M. Zeiler and R. Fergus. Visualizing and understanding convolutional networks. *European Conference on Computer Vision (ECCV)*, 8689:818–833, 2014. 2

[28] R. Zhang, P. Isola, and A. A. Efros. Colorful Image Colorization. In *European Conference on Computer Vision (ECCV)*, 2016. 1, 2, 3, 4, 5, 6, 7, 8