**SRM: A Style-based Recalibration Module for Convolutional Neural Networks**Adi Horowitz, Gilad Deutch

1. **Abstract**

*In our project paper, we aim to improve the representational ability of a CNN, by considering the style of the image, as defined below. SRM stands for a Style-based Recalibration Module. This module is directly fed into existing CNN architectures. By incorporating the relative importance of individual styles into feature maps, SRM learns to suppress the contribution of nuisance styles, and by doing that, the network can concentrate more on meaningful features, which boost its representational power. Finally, we added some improvements of our own. We examined the option of adding new features to expand the style features vector. Later on, we tested the importance of the SRM layer location inside the network. We performed experiments on a few general vision tasks and show the improvement of results when using SRM over other approaches such as Squeeze-and-Excitation (SE). We also show the results of SRM with our own modifications and additions. Throughout the experiments, SRM outperforms recent approaches though it requires orders of magnitude less additional parameters.*

1. **Intro**

It has been known that CNNs are capable of handling not only the content (i.e. shape) but also the style (i.e. texture) of an image. The distinction between the two makes it possible to not only perform style transfer, but improve any kind of vision recognition problem. We will start by mentioning some relative work. Gatys et al. discovered that the feature statistics of a CNN effectively encode the style information of an image. Based on that discovery, our project paper writers used mean and standard division statistics in order to capture the style of the input image, as we will explain later in the methods section. This discovery also laid the foundation for neural style transfer, hence it was important for many further researches. Adaptive instance normalization (i.e. AdaIN) further showed that transferring channel-wise mean and standard deviation can efﬁciently change image styles. Later on, Karras et al. combined AdaIN into generative adversarial networks (GANs) to improve the generator by adjusting styles in intermediate layers. Geirhos et al. discovered that CNNs are highly biased towards styles in their decision making process. Recent papers (e.g. Brendel and Bethge) pointed out as well that the style information of an image has a significant role in the decision making process done by standard CNNs. This could be the motive for our project paper, which tries to take advantage of the style information in a CNN, in form of features recalibration based on the style. The style-based-recalibration module is composed on a standard CNN, in the purpose of making the most out of the image style information, in order to improve performance. In general, this is done by recalibrating the feature map according to meaningful style features. We also study Gatys et al. who presented impressive style transfer results by exploiting the second-order statistics (i.e. the Gram matrix) of convolutional features as style representations. One of our ideas for modifications is based on that article. We attempt to integrate the gram matrix into the style feature vector used in the style-recalibration module.

More related to our work, Squeeze-and-Excitation (SE) proposed a channel-wise recalibration operator that incorporates the interaction between channels. Similar to SRM module, SE consist of two components. It ﬁrst aggregates the spatial information with global average pooling, and then captures the channel dependencies using a fully connected subnetwork. However, in contrast to the prior efforts, our project paper reformulate channel-wise recalibration in terms of leveraging style information, without the aid of channel relationship nor spatial attention. We present a style pooling approach which is superior to the standard global average used in the SE recalibration method. Moreover, our project paper introduce a channel-independent style integration method which is substantially more lightweight than fully connected, used in SE blocks as well. SRM is designed to be light weight in both terms of memory and computational complexity, yet more effective in various scenarios. Despite its minimal overhead, SRM shows an improvement while comparing it to the other approach.

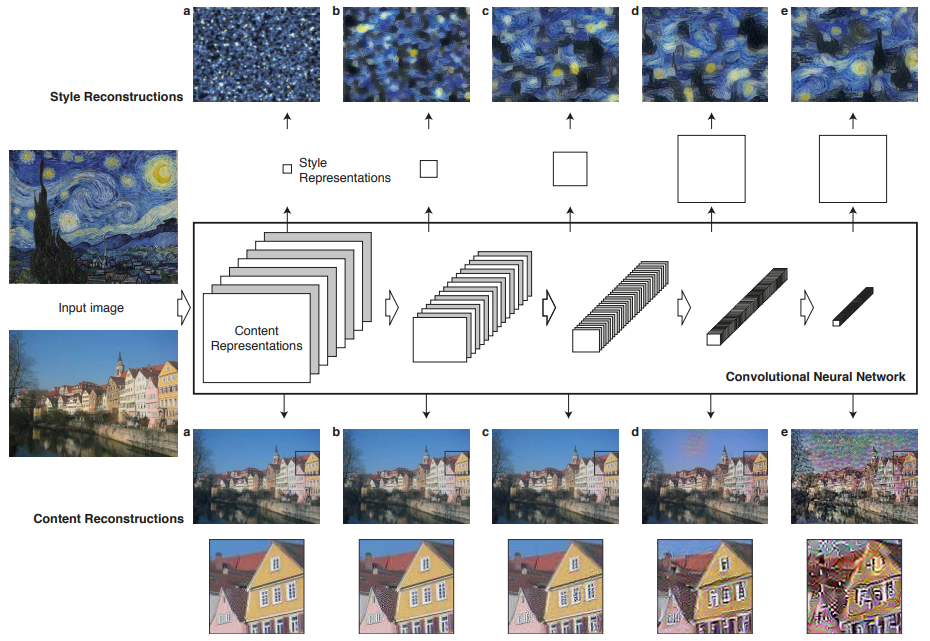
We aim to enhance the utilization of styles in a CNN to boost its representational power, in order to leverage the capabilities of CNN network with a wide range of vision classification problems, While only impose negligible additional parameters and computations.

1. **Methods**

As mentioned before, the proposed recalibration module consists of two main components: style pooling, and style integration. SRM ﬁrst extracts the style information from each channel of the feature maps by style pooling. Then, estimates per-channel recalibration weight via channel-independent style integration. In our implementation, we used convolution layer with kernel size equals the number of style features. In this case the CNN essentially becomes fully connected channel-independent layer, which is needed in the integration stage.

In the paper, the writers focused on the channel-wise statistics for efﬁciency and conceptual clarity. Therefore, the style features vector in the article contain only two elements: the mean, and the standard division. We have examine the module when adding more features into the style features vector. As part of our attempts for improving the original paper, we tried to include the correlation between different channels. The theoretical motivation for it is that by including the feature correlations of multiple layers, we obtain a stationary, multi-scale representation of the input image, which captures its texture information but not the global arrangement. Another modification to the original SRM model that we tried to experiment with is using the median. As mentioned, in the original paper an image style was defined as the average and std of each of its channels. Intuitively, we can see that this definition is a rough approximation of the image distribution. In order to better represent the image's distribution, we also used the median in addition to the mean and std. Furthermore, we combined the approach using the feature correlations, and the approach using the median to create a SRM model that uses both the median and the feature correlations (in addition to the mean and std, of course). Moreover, we try to apply the recalibration module in different stages of the CNN representation. Our assumption, based on Figure 1, is that this kind of recalibration module would be more efficient when applying it into a low-level features representation. Nevertheless, we experimented with the location of the SRM module in the neural net in order to approve this hypothesis.

The data we are going to use for evaluating is CIFAR10 database.



**Figure 1.** taken from [4]. A given input image is represented as a set of filtered images at each processing stage in the CNN. While the number of different filters increases along the processing hierarchy, the size of the filtered images is reduced by some downsampling mechanism (e.g. max-pooling) leading to a decrease in the total number of units per layer of the network. Content Reconstructions. We can visualise the information at different processing stages in the CNN by reconstructing the input image from only knowing the network’s responses in a particular layer. **The style is most pronounced at low-level features representation.**

1. **Implementation and experiments**

In order to verify the effectiveness of SRM, we conduct several experiments using different general object classiﬁcation problems and datasets.

First, we evaluate the performance of SRM on the CIFAR-10 dataset, in comparison with the state-of-the-art method Squeeze-and Excitation (SE).

Later on, we also evaluate the performance of SRM on the ImageNet-1K.

We implemented all competitors to compare under consistent settings for fair comparison.

1. **Results**

Throughout the experiment, SRM outperforms recent approaches though it requires orders of magnitude less additional parameters.

boosting meaningful features while suppressing weak ones,

While certain styles play an essential role, some are rather a nuisance factor to the task [25]. SRM dynamically estimates the relative importance of individual styles then reweights the feature maps based on the style importance, which allows the network to focus on meaningful styles while ignoring unnecessary ones.

**References –**

[1] [Recalibrating Fully Convolutional Networks with Spatial and Channel ‘Squeeze & Excitation’ Blocks](https://arxiv.org/pdf/1808.08127.pdf)

**[2]**  [P-CNN: Part-Based Convolutional Neural Networks for Fine-Grained Visual Categorization](https://ieeexplore.ieee.org/document/8789527)

# [3] [Visualising Filters and Feature Maps for Deep Learning](https://towardsdatascience.com/visualising-filters-and-feature-maps-for-deep-learning-d814e13bd671)

# [4] [Image Style Transfer Using Convolutional Neural Networks](https://zpascal.net/cvpr2016/Gatys_Image_Style_Transfer_CVPR_2016_paper.pdf)

[5] L. A. Gatys, A. S. Ecker, and M. Bethge. Texture synthesis using convolutional neural networks. In NIPS, 2015. 1