Deep Learning for Paint Loss Detection: A Case Study on the Ghent Altarpiece

Laurens Meeus[1], Shaoguang Huang[1], Bart Devolder[2], Maximiliaan Martens[3] and Aleksandra Pižurica[1]

[1]TELIN-IPI, Ghent University - imec, Belgium

[2]Princeton University Art Museum, United States of America.

[3]Department of Art, Music and Theatre Sciences, Ghent University, Belgium.

***Abstract*—Producing damage surveys as part of condition reports prior to and during restoration treatments is often a tedious and time-consuming work for the art restorer. We explore the potential of deep learning for automatic paint loss detection in paintings to facilitate condition reporting and to support restoration treatments. To the best of our knowledge, this is the first reported attempt of employing deep learning in this application. We develop a multiscale deep learning method, based on the recent U-Net architecture which we extend with dilated convolutions, such as to improve the detection stability. Our model is applicable to multimodal acquisitions such as visible, infrared, x-ray, and ultraviolet fluorescence. As a case study we use multimodal data of the *Ghent Altarpiece*. Our results indicate huge potential of the proposed approach in terms of accuracy and also its exceptional speed which allows interactivity and continuous learning.**

1 Introduction

One of the documentation tasks during the conservation/restoration of paintings consists of mapping lacunas as well as larger paint losses. Lacunas are mostly a result of drying and flaking of paint, although rough handling can also introduce losses. Currently, the mapping involves a lot of manual work since available software can only give a coarse estimation of the paint loss. This makes the process rather slow and tedious. In order to improve the automated mapping, smarter image processing techniques are sought.

Paintings are nowadays typically scanned in different modalities prior to restoration treatments and during their various stages. Hence, our approach will be designed to make use of the multimodal data. As the size of losses can range from a few to hundreds of pixels, the algorithm should not only take into account spectral information, but also have a large enough spatial support.

Technical literature on paint loss detection is limited. Huang et al [1] reported promising results with sparse representation classification (SRC), surpassing common machine learning approaches like linear regression classification and support vector machines in this task. We propose an alternative method based on deep learning, motivated by the huge success of convolutional neural networks in many other image classification and segmentation problems. We will validate our method on the panels of the *Ghent Altarpiece* [2], a monumental triptych made by the brothers van Eyck in the 15^{th} century. To the best of our knowledge we are the first to report a deep learning method for paint loss detection.

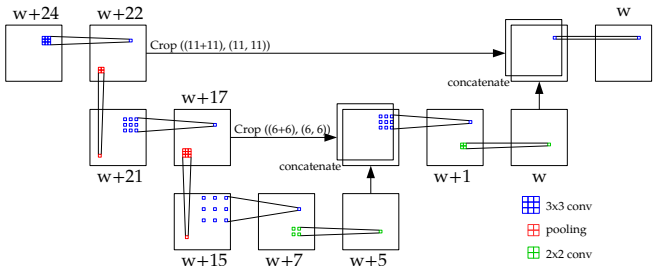
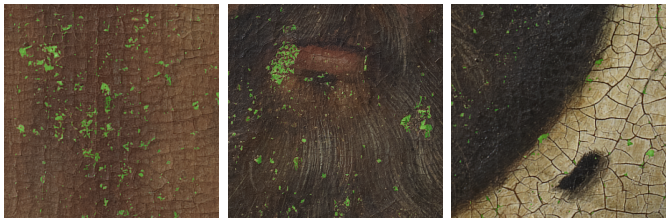
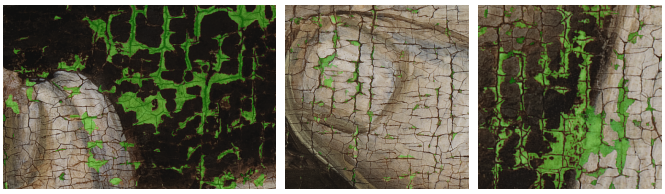


Figure 1: Proposed network architecture: a multiscale, multipath network with dilated convolutions.



(a) Details from panel *Prophet Zachary*.



(b) Details from panel *John the Evangelist*.

Figure 2: Annotations made by the art restorer. Craquelure, formed by ageing, is not considered paint loss and is treated differently from it.

2 Methods

The proposed neural network architecture is visualised in figure 1. Similar to the U-Net [3] it consists of an encoder (left), a decoder (right) and skip-connect layers between the encoder and decoder (top) [4]. The difference between the U-Net and the proposed architecture is the removal of the decimation in the pooling layers. This way we maintain the same resolution in all layers and enforce true translation invariance. While this makes the bottom layers more dense than in the original U-Net, the outputs become more averaged out and this improves the stability of the output values. We observe that this leads to an increase in accuracy and learning capability of the model. The encoder consists of 3×3 convolutional layers and for the activation function the Rectified Linear Unit (ReLU, $\sigma(z) = \max(0, z)$) is used. Between these layers, pooling is introduced by taking the maximum in a 2×2 window with overlap to maintain the resolution. To maintain the same receptive field, the subsequent layers are replaced with dilated convolu-

(a) Original image during treatment. (b) Detected paint loss. (c) Original image during treatment. (d) Detected paint loss.

Figure 3: Paint loss detection on parts of the grisaille panel: *John the Evangelist*. The following modalities are provided to the model: visible before restoration, visible after varnish and over-paint removal, and infrared.

tions [5] and the amount of kernel weights remains identical with respect to the original U-Net. The decoder mirrors the encoder and pooling layers are replaced with upsampling 2×2 convolutional layers with linear activation. The skip-connect copies the layer of the encoder and concatenates it with the output of the upsampling layer to combine information of layers working at different resolutions. This gives the network the possibility to learn features on multiple scales simultaneously along the different paths. The last layer is a per-pixel, fully connected layer producing 2 feature maps: the probabilities of a pixel being paint loss or not. These probabilities are a result of the non-linear activation function Softmax:

$$\sigma(\mathbf{z})_j = \frac{e^{z_j}}{\sum_{k=1}^{2} e^{z_k}}. \tag{1}$$

It converts each pixel values $\mathbf{z}$ to a normalised probability vector $\hat{\mathbf{y}} = [p_0, p_1], p_0 + p_1 = 1)$.

To train the filters of the CNN, annotated data is requested. In our case, these were provided by art restorers of the *Ghent Altarpiece*. For each pixel, the annotation is converted to a vector $\mathbf{y}_i = [1; 0]$ for not paint loss and $\mathbf{y}_i = [0; 1]$ for paint loss. The CNN is trained to minimise the cross entropy:

$$C = -y_0 \cdot \log \hat{y}_0 - y_1 \cdot \log \hat{y}_1 \tag{2}$$

using Adaptive Moment Estimation [6]. The final prediction map is obtained by thresholding the probability p_1 of the output. We obtained the highest Intersection over Union score by thresholding at 0.5.

For the input of the network, the different modalities are first registered, concatenated and then cropped to a fixed size. Since each convolution and pooling operation reduces the output area, the input patch of the network is larger than the output patch to account for the receptive field. Because the input shape is a fixed amount bigger than the output shape and all layers operate at the same resolution, there is freedom in selecting the size of the patch to be segmented. Instead of classifying each pixel individually by setting the output shape to 1×1, it is more efficient to classify a big patch of pixels at once. When classifying nearby pixels, the overlap of the receptive field allows the convolutional layers to share computations. This speeds up the inference significantly and this means for the end user a big difference for practical usage.

3 Results and discussion

Figure 3 visualises the detection results on a larger part of the panel *John the Evangelist*. The 6 regions of the *Ghent Altarpiece* annotated by the art restorer, illustrated in figure 2, are from the panels *Prophet Zachary* and *John the Evangelist*. In total there are $807,740$ annotated pixels available of which 8.3% is paint loss. This amount is increased by a factor 8 after data augmentation by rotations of $90°$ and flips. These annotated regions are divided into smaller patches after which the network is trained on 80% of these patches. The remainder is used for picking the optimal hyperparameters and testing the accuracy. The following modalities were given to the model: optical images before and during treatment, infrared, infrared reflectography, X-ray, and ultraviolet fluorescence.

By segmenting patches of 10×10 or 100×100 instead of per pixel, we observe a speed increase of a factor 40 and 300 respectively for the inference. The results in figure 3 illustrate the binary prediction of a relatively large image (size 5954×7545), processed in less than a minute on a GeForce GTX 1070. Our experiments indicate a stable performance even with relatively few annotations. While our technique achieves similar results as the SRC-based method of [1], it is orders of magnitude faster. The art restorers appreciate the achieved results and the speed shows a huge potential for practical use of the proposed approach.

References

[1] S. Huang, W. Liao, H. Zhang, and A. Pižurica, "Paint loss detection in old paintings by sparse representation classification," in *International Traveling Workshop On Interactions Between Sparse Models And Technology*, (Aalborg, Denmark), pp. 62–64, August 2016.

[2] KIK/IRPA, "Belgian art links and tools," 2018.

[3] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," *ArXiv e-prints*, May 2015.

[4] M. Drozdzal, E. Vorontsov, G. Chartrand, S. Kadoury, and C. Pal, "The importance of skip connections in biomedical

image segmentation," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2016.

[5] F. Yu and V. Koltun, "Multi-Scale Context Aggregation by Dilated Convolutions," *ArXiv e-prints*, November 2015.

[6] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," pp. 1–15, 2014.