

AI classifier of defects in Artworks captured by active infrared thermography

Peter Malík*, Martin Orlej†, Branislav Fajčák† and Massimo Rippa§

*Institute of Informatics, Slovak Academy of Sciences,
Dúbravská cesta 9, 84507 Bratislava, Slovakia
Email: p.malik@savba.sk, ID:0000-0002-1921-2340

†Faculty of Informatics and Information Technologies, Slovak University of Technology
Ilkovičova 2, 84216 Bratislava, Slovakia
Email: xorlej@stuba.sk, fajcak.branislav@gmail.com

‡Institute of Applied Sciences and Intelligent Systems “Eduardo Caianiello”, National Research Council,
Via Campi Flegrei 34, Pozzuoli, 80078 Naples, Italy
Email: massimo.rippa@isasi.cnr.it, ID:0000-0002-1993-4589

Abstract—Automatic recognition of features in digital images has become a central topic in the field of cultural heritage diagnostics. AI-based models are being increasingly applied to the analysis of infrared reflectography and thermographic data. They show great promise in automating time-consuming manual analyses and improving the objectivity and repeatability of diagnostic assessments. This work proposes 4 specialized classifiers for nails and detachments in work of arts. In-situ active thermography measurements are used for training proposed models. AI models for nail classification reached accuracy of 96.03 % and 93.65 % using planar composite thermal images and volumetric raw data as inputs, respectively. AI models for detachment classification reached accuracy of 87 % and 57 % using planar composite thermal images and volumetric raw data as inputs, respectively.

I. INTRODUCTION

IN RECENT years, the automatic recognition of features in digital images has become a central topic in the field of cultural heritage diagnostics, particularly with the increasing adoption of imaging-based techniques. Advances in computer vision and deep learning have led to the development of algorithms capable of identifying patterns, anomalies, and structural details that are often imperceptible to the human eye. Artificial intelligence-assisted visual inspection for cultural heritage is reviewed in [1]. Work [2] discusses application of artificial intelligence in cultural heritage protection. Pattern recognition and artificial intelligence techniques for cultural heritage are analyzed in [3]. Detecting treasures in museums with artificial intelligence is described in [4]. Reducing bias in AI-based analysis of visual artworks is discussed in [5]. Work [6] analyzes the research trends for using artificial intelligence in cultural heritage. Convolutional neural networks (CNNs) and other AI-based models are being increasingly applied to the analysis of infrared reflectography and thermographic data, with the potential to support tasks such as

detecting underdrawings, material heterogeneities, and traces of previous restoration interventions. These approaches show great promise in automating time-consuming manual analyses and improving the objectivity and repeatability of diagnostic assessments. Nonetheless, challenges remain regarding the availability of training data, the need for domain-specific model tuning, and the effective integration of these tools into conservation workflows. More scientific works are dedicated to artwork inspection. Physical degradation detection on artwork surface polychromies using deep learning models is discussed in [7]. A deep learning approach for anomaly detection in X-ray images of paintings is described in [8].

This work presents 4 specialized classifiers for nails and detachments in work of arts. Dataset for training AI models was created from in-situ active thermography measurements. Quantity of measurement is low which makes the process to design AI models more difficult. AI models for nail classification reached high accuracies of 96.03 % and 93.65 % for models optimized for planar composite thermal images and volumetric raw data measurements, respectively. AI models for detachment classification reached accuracies of 87 % and 57 % for models optimized for planar composite thermal images and volumetric raw data measurements, respectively. Main contribution of this work is in created dataset and 4 specialized AI classifiers.

Proposed models are designed for use before restoration process or for digitization purposes. They are created as useful tool that reduces long manual work required for analysis of a large size work of Art. They can be additionally use or making the process of artwork analysis available for less experienced personnel.

II. DATA PREPARATION

Active Thermography (AT) measurements were carried out following a standardized procedure for each selected area of the paintings. A thermal stimulus lasting 10 seconds was applied using a 1000 W halogen lamp with adjustable output. Depending on the characteristics of each painting, the

This work was supported by the Slovak Scientific Grand Agency VEGA under the contract 2/0135/23 "Intelligent sensor systems and data processing". Funded by the EU NextGenerationEU through the Recovery and Resilience Plan for Slovakia under the project No. 09I05-03-V02-00055.



Fig. 1. Artwork example - Adorazione dei Magi painting, author M. Cardisco, size 254 x 268 cm

applied power ranged between 200 and 300 W. The lamp was positioned approximately 1–2 meters from the surface. A FLIR X6580 sc infrared camera (equipped with a cooled InSb detector, operating in the MWIR range of 3.5–5 μm , IFOV 0.3 mrad, NETD 20 mK at 25 μm , and a 640 $\times$ 512 FPA sensor) with a 50 mm focal length lens was employed to record thermal sequences.

Temperature rise during heating was controlled in real-time using the ResearchIR software (FLIR Systems Inc., Wilsonville, OR, USA), ensuring that the maximum temperature difference (ΔT) on the surface did not exceed 5 $^{\circ}\text{C}$ and was as uniform as possible. After the thermal pulse, sequences of 150 to 300 thermal images were acquired for each area at a frame rate of 5 Hz. The data were exported in CSV format and subsequently analyzed using two established post-processing techniques: Principal Component Thermography (PCT) and Thermal Recovery Mapping (TRM), both well-documented in the literature [9], [10], [11], [12].

Custom-developed scripts in MATLAB (R2019a, MathWorks) were used for data processing, which was performed on a workstation equipped with an Intel i7-4770 CPU @ 3.40 GHz (8 cores) and 32 GB RAM. These analyses enabled (a) the identification and enhancement of major thermal anomalies within 1–2 Spatial Components (SCs) through PCT, and (b) their classification, with particular attention to detecting detachments and metallic inclusions (e.g., nails). The thermal datasets corresponding to these two types of defects were subsequently organized into collections of raw thermal images and their associated SCs, which served as the basis for developing the analytical models proposed in this study. Altogether, 4 paintings have been processed, see TABLE I, TABLE II and Fig. 1. Due to larger size of paintings, each painting was captured and evaluated by grid sections resulting in 18 sections with 76 labeled nails and 15 sections with 30 labeled detachments, see Fig. 2 and Fig. 3.

TABLE I
SOURCES AND STATISTICS OF NAIL DEFECTS

Author	Artwork Name	Nails	Related samples
M. Cardisco	Adorazione dei Magi	8	N1-N3
Unknown	Eternal Father	3	N4
Unknown	Madonna del Pozzano	18	N5-N7
Unknown	San Francesco	47	N8-N18

TABLE II
SOURCES AND STATISTICS OF DETACHMENT DEFECTS

Author	Artwork Name	Detachments	Related samples
M. Cardisco	Adorazione dei Magi	17	D1-D9
Unknown	Eternal Father	13	D10-D15
Unknown	Madonna del Pozzano	0	n.a.
Unknown	San Francesco	0	n.a.

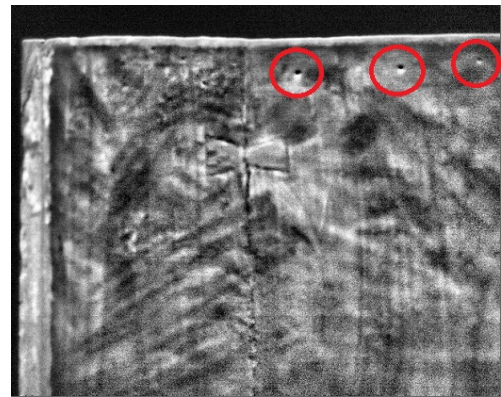


Fig. 2. An SCs section example with labeled nails, Adorazione dei Magi authored by M. Cardisco



Fig. 3. An SCs section example with labeled detachments, Adorazione dei Magi authored by M. Cardisco

TABLE III
STATISTICS OF CREATED DATASETS

Dataset name	Number of samples	Spatial resolution	Number of channels
Nails SCs	626	64 x 64	1
Nails raw data	626	64 x 64	300
Detachments SCs	114	180 x 175 (on average)	1
Detachments raw data	114	180 x 175 (on average)	300

A. Dataset preparation

Due to low quantity of nail and detachment examples, the binary classification of nails and detachments was selected as the best task to implement. Multi class classification and detection are more complex and usually requires more parameters to train. Our very small dataset limits us to train smaller models. Bigger models with more parameters require larger datasets [13], [14] to be sufficiently trained. Two specialized binary classifiers allow to focus on each problem characteristics more deeply and maximized prediction accuracy with less data.

Dataset is created with cropped areas. We experimented with various input resolutions. For nails, 64 x 64 pixels was selected. With higher resolutions, accuracy is decreasing significantly. Smaller resolution offers slightly higher accuracy but may be prone to overspecialization. For detachments, rectangles with longer size 180 pixels was selected. The dataset is balanced with 50 % positives and 50 % negatives (background). Focus was on including hard negatives with very close visual resemblance to positives. Min-max normalization was applied to raw data (CSVs) to improve learning. Missing channels were padded with the last measured channel. This way we don't introduce any bias by padding the data with either white, black or gray color equivalent. The quantity was increased with augmentation techniques [15] to partially alleviate low quantity of data and to improve the trained models' robustness. Statistics are shown in TABLE III.

III. ARCHITECTURE DESIGN

Short visual analysis revealed that variation of detachment examples is much more varied which makes it significantly more difficult task to classify in comparison to nails. The number of detachment true examples is only 30 which represents around 40 % of quantity of true nail examples and such low quantity makes the problem even more difficult. We started our design approach with SCs images due to its simpler planar form. Designing a model for volumetric raw data is more difficult. The dataset was split 80 % – 20 % for training and testing. Presented values in following Tables are averaged results of several training runs started from scratch with random seeding.

TABLE IV
VARIOUS 2D CONVOLUTION MODELS

Model depth	Convolution kernels	Pooling kernels	Validation Accuracy	Test Accuracy
1	(7,7)	Glob.	70.20 %	80.95 %
1	(11,11)	Glob.	85.88 %	82.54 %
2	(5,5);(5,5)	(4,4);Glob.	81.96 %	87.30 %
2	(7,7);(7,7)	(4,4);Glob.	75.29 %	75.13 %
2	(3,3);(3,3)	(6,6);Glob.	86.47 %	90.48 %
2	(5,5);(3,3)	(6,6);Glob.	90.00 %	90.48 %
3	(3,3);(3,3);(3,3)	(2,2);(2,2);Glob.	91.76 %	88.89 %
3	(3,3);(3,3);(3,3)	(2,2);(4,4);Glob.	82.35 %	82.54 %
3	(5,5);(5,5);(5,5)	(2,2);(4,4);Glob.	47.45 %	57.14 %
3	(3,3);(3,3);(2,2)	(3,3);(3,3);Glob.	86.47 %	90.48 %
3	(3,3);(2,2);(2,2)	(3,3);(3,3);Glob.	86.47 %	85.72 %
4	(3,3);(3,3); (3,3);(3,3)	(2,2);(2,2); (2,2);Glob.	68.25 %	69.31 %

TABLE V
EFFECTS OF GLOBAL POOLING AND EXTENDED CLASSIFIER

Global pooling	Added layers	Validation Accuracy	Test Accuracy
Global Max Pool	none	81.96 %	87.30 %
Global Max Pool	2D conv. (1,1) x64	76.93 %	75.40 %
Global Average Pool	none	69.41 %	69.84 %
Global Average Pool	2D conv. (1,1) x64	68.12 %	73.86 %

A. Nail defect optimized models using SCs images

The training setup for nail classification using SCs images is following: Environment TensorFlow, Adam optimizer, 20 Epoch training length, LR started from 0.0008, LR minimal set to 0.0001, LR reducing factor set to 0.5, input resolution 64 x 64 pixels, quantity of samples 626. Adam optimizer provided the best results.

We experimented with different augmentations. The standard rotations and flipping are already included in dataset. Additive noise hurt accuracy with pink noise being worse in comparison to white noise. Varied Gamma, implemented by a constant additive shift, produced varied and conflicted results. We decided not to use it due to unstable improvements and very small increases.

Experiments with various model architectures are shown in TABLE IV. Results show that the optimal model depth is 2 or 3. The single layer models are not descriptive enough and 4 layer deep model is probably too complex with many parameters for this dataset. Smaller convolution kernel sizes are beneficial for 2 and 3 layer deep models. The 2 layer deep model reached the best accuracy and was selected. Variation of using different global pooling and extensions can be seen in TABLE V. The global Max pooling is the best option.

The effects of different width of layers (varied number of filters) can be seen in TABLE VI. Decreasing or increasing number of filters symmetrically in all layers has little effect with 64 filters being the best. Results of additional architecture

TABLE VI
EFFECTS OF LAYERS' WIDTH

Number of filters	Validation Accuracy	Test Accuracy
32	69.28 %	68.37 %
64	69.41 %	69.84 %
96	68.77 %	68.51 %
128	67.96 %	68.91 %

TABLE VII
VARIOUS 2D CONVOLUTION MODELS, CONTINUATION

Model depth	Convolution kernels	Pooling kernels	Validation Accuracy	Test Accuracy
2	(7,7)x64;(3,3)x64	(6,6);Glob.	90.00 %	88.89 %
2	(9,9)x64;(3,3)x64	(6,6);Glob.	94.71 %	92.86 %
2	(5,5)x64;(3,3)x64	(8,8);Glob.	87.06 %	92.86 %
2	(7,7)x64;(3,3)x64	(8,8);Glob.	93.53 %	91.27 %
2	(9,9)x64;(3,3)x64	(8,8);Glob.	92.35 %	92.06 %
2	(7,7)x128;(3,3)x64	(6,6);Glob.	91.18 %	92.86 %
2	(9,9)x128;(3,3)x64	(6,6);Glob.	90.59 %	89.68 %
2	(5,5)x128;(3,3)x64	(8,8);Glob.	84.12 %	88.89 %
2	(7,7)x128;(3,3)x64	(8,8);Glob.	90.59 %	87.30 %
2	(9,9)x128;(3,3)x64	(8,8);Glob.	94.71 %	96.03 %
2	(7,7)x256;(3,3)x64	(6,6);Glob.	88.82 %	88.10 %
2	(9,9)x256;(3,3)x64	(6,6);Glob.	90.00 %	89.68 %
2	(5,5)x256;(3,3)x64	(8,8);Glob.	83.53 %	89.68 %
2	(7,7)x256;(3,3)x64	(8,8);Glob.	91.18 %	91.27 %
2	(9,9)x256;(3,3)x64	(8,8);Glob.	93.53 %	94.45 %

search is shown in TABLE VII. Larger kernels of the first layer have positive effect. The best model has 128 filters with kernel size of 9 x 9 in the first layers and Max pooling of 8 x 8.

B. Nail defect optimized models using raw data

The training setup for nail classification using raw data measurements is following: Environment TensorFlow, Adam optimizer, 20 Epoch training length, LR started from 0.0003, LR minimal set to 0.00002, LR reducing factor set to 0.5, input resolution 64 x 64 pixels, quantity of samples 626.

Experiments with various depth of model architecture are shown in TABLE VIII. All models use 3D convolutions and 3D Max pooling. Each convolution layer is followed by ReLU [16] nonlinearity and Max pooling. The optimal model depth is around 3. Resulted accuracy is lower in comparison to models using SCs images. We focused on 3 layer deep architectures, see TABLE IX. Results show that the bigger kernels have positive effect.

The small dataset is a very limited factor for volumetric data. Additional augmentations fail to improve results. So we opted to reduce input space by Principal Component Analysis (PCA). We applied PCA to dataset and used results as inputs, see TABLE X. The best accuracy was reached by using 50 most significant PCA components.

We continued with architecture exploration with 50 PCA components as inputs, see TABLE XI. Results shows that the

TABLE VIII
VARIOUS 3D CONVOLUTION MODELS

Model depth	Convolution kernels	Pooling kernels	Validation Accuracy	Test Accuracy
1	(3,3,3)	Glob.	64.71 %	61.90 %
2	(3,3,3);(3,3,3)	(2,2,2);Glob.	71.76 %	65.08 %
3	(3,3,3);(3,3,3) (3,3,3)	(2,2,2);(2,2,2); Glob.	75.29 %	71.43 %
4	(3,3,3);(3,3,3) (3,3,3);(3,3,3)	(2,2,2);(2,2,2); (2,2,2);Glob.	75.29 %	66.67 %
5	(3,3,3);(3,3,3) (3,3,3);(3,3,3) (3,3,3)	(2,2,2);(2,2,2); (2,2,2);(2,2,2); Glob.	67.06 %	57.14 %

TABLE IX
VARIOUS 3D CONVOLUTION MODELS

Model depth	Convolution kernels	Pooling kernels	Validation Accuracy	Test Accuracy
3	(3,3,3);(3,3,3) (3,3,3)	(2,2,2);(2,2,2); Glob.	75.29 %	71.43 %
3	(1,3,3);(3,1,1); (3,3,3)	(2,2,2);(2,2,2); Glob.	75.29 %	63.49 %
3	(5,1,1);(1,5,5); (3,3,3)	(2,2,2);(2,2,2); Glob.	72.94 %	60.32 %
3	(5,3,3);(3,5,5); (3,3,3)	(2,2,2);(2,2,2); Glob.	77.65 %	73.02 %
3	(3,5,5);(5,3,3); (2,2,2)	(2,2,2);(2,2,2); Glob.	81.18 %	76.19 %

suitable depth is at least 3. Smaller convolution kernels have positive effect, especially at later layers. Smaller input map before the last global pooling is beneficial. This is especially true with the deepest model which has the best test accuracy. Higher parameter count is detrimental with very small datasets which makes the best model with higher parameter count a surprise. It is interesting that using smaller convolution kernel than (3,3,3) anywhere in architecture reduces the accuracy. However, the optimal trends are clouded by the small dataset size and it can be seen as fluctuating results. We experimented with the varied width of layers with 64 filters being the best.

TABLE X
EFFECTS OF USING PCA TO REDUCE NUMBER OF INPUT CHANNELS

Number of PCA components	Validation Accuracy	Test Accuracy
300 (no PCA)	84.71 %	82.54 %
200	84.71 %	80.95 %
150	85.88 %	82.54 %
50	94.12 %	90.48 %
25	90.48 %	85.71 %
15	85.88 %	79.37 %
10	83.54 %	80.95 %

TABLE XI
VARIOUS 3D CONVOLUTION MODELS

Model depth	Convolution kernels	Pooling kernels	Validation Accuracy	Test Accuracy
2	(5,5,5);(5,5,5)	(5,5,5);Glob.	94.12 %	87.30 %
3	(5,5,5);(3,3,3); (3,3,3)	(3,3,3);(3,3,3); Glob.	91.76 %	92.06 %
3	(5,5,5);(3,3,3); (3,3,3)	(3,3,3);(2,2,2); Glob.	89.41 %	88.89 %
3	(3,5,5);(5,3,3); (2,2,2)	(2,2,2);(2,2,2); Glob.	94.12 %	90.48 %
4	(3,3,3);(3,3,3); (3,3,3);(3,3,3)	(2,2,2);(2,2,2); (2,2,2);Glob.	88.24 %	92.06 %
5	(3,3,3);(3,3,3); (3,3,3);(3,3,3); (3,3,3)	(2,2,2);(2,2,2); (2,2,2);(2,2,2); Glob.	91.76 %	93.65 %
3	(3,3,3);(3,3,3); (2,2,2)	(3,3,3);(3,3,3); Glob.	89.42 %	81.75 %
3	(2,2,2);(2,2,2); (2,2,2)	(3,3,3);(3,3,3); Glob.	85.88 %	80.95 %
4	(3,3,3);(3,3,3); (2,2,2);(2,2,2)	(2,2,2);(2,2,2); (2,2,2);Glob.	90.59 %	80.95 %
4	(2,2,2);(2,2,2); (2,2,2);(2,2,2)	(2,2,2);(2,2,2); (2,2,2);Glob.	89.41 %	88.89 %
5	(3,3,3);(3,3,3); (3,3,3);(2,2,2); (2,2,2)	(2,2,2);(2,2,2); (2,2,2);(2,2,2); Glob.	90.59 %	90.48 %
5	(3,3,3);(2,2,2); (2,2,2);(2,2,2); (2,2,2)	(2,2,2);(2,2,2); (2,2,2);(2,2,2); Glob.	94.12 %	87.30 %
5	(2,2,2);(2,2,2); (2,2,2);(2,2,2); (2,2,2)	(2,2,2);(2,2,2); (2,2,2);(2,2,2); Glob.	90.59 %	92.06 %

C. Detachment defect optimized models using SCs images

The training setup for detachment classification using SCs images is following: Environment TensorFlow, Adam optimizer, 50 Epoch training length, LR started from 0.0001, LR minimal set to 0.000001, LR reducing factor set to 0.5, input resolution 115 x 115 & 128 x 128 & 130 x 130 & 150 x 150 pixels except where it is noted, augmentations: LR flips & 90° angle rotations & brightness adjustments (gamma variations). One transformation is randomly applied to each image. These augmentations were applied only to the train and validation datasets during training.

We experimented with various scaling of the inputs. The best performance was achieved with the input resizing to 128 x 128 pixels resolution using stretching, see TABLE XII. We augmented data with various input sizes which significantly improved accuracy. Our experiments with different padding schemes confirmed that the stretching is the best, see TABLE XIII.

Experiments with various model architecture are shown in TABLE XIV. The optimal model depth is 2. However, 4 layer deep model reached good accuracy. Optimal kernel size

TABLE XII
EFFECTS OF VARIOUS INPUT RESOLUTIONS

Input size configuration	Validation Accuracy	Test Accuracy
128 x 128 (Resized only)	71 %	79 %
Original sizes	61 %	60 %
Original sizes and 128 x 128	71 %	77 %
128 x 128 & 115 x 115 & 150 x 150 & 130 x 130	78 %	87 %

TABLE XIII
PADDING VARIATIONS FOR CREATING INPUT SAMPLES FOR DATASET

Padding type	Validation Accuracy	Test Accuracy
Zero padding	61 %	48 %
Mean value padding	65 %	61 %
Mirror padding	65 %	63 %
Stretching	78 %	87 %
No padding	61 %	60 %
Up to 10 % mirroring padding	69 %	77 %
Up to 20 % mirroring padding	67 %	72 %
Up to 30 % mirroring padding	65 %	69 %

varies. More experiments are necessary. Our experiments with different global pooling confirmed the Max pooling as the best, see TABLE XV. The global average pooling with adding an extra layer behind it equaled the global Max pooling. We selected global Max pooling due to lower parameter count. We experimented with the different width, but for most models, 64 filters is the best option. Our best model uses small kernels and twice the number of filters in the first layer. We also experimented with different additional augmentations. Varied contrast have negative effect. Surprisingly, added small amount of white or pink noise have small positive effect.

TABLE XIV
VARIOUS 2D CONVOLUTION MODELS

Model depth	Convolution kernels	Pooling kernels	Validation Accuracy	Test Accuracy
1	(15,15)	Glob.	66 %	68 %
1	(25,25)	Glob.	67 %	66 %
2	(3,3)x256;(3,3)x128	(2,2);Glob.	77 %	87 %
2	(11,11);(11,11)	(4,4);Glob.	75 %	83 %
2	(15,15);(15,15)	(5,5);Glob.	71 %	81 %
3	(7,7);(7,7); (7,7)	(4,4);(4,4); Glob.	69 %	67 %
3	(11,11);(11,11); (9,9)	(4,4);(4,4); Glob.	67 %	59 %
4	(5,5);(5,5); (5,5);(3,3)	(3,3);(3,3); (4,4);Glob.	77 %	70 %
4	(7,7);(7,7); (7,7);(5,5)	(3,3);(3,3); (4,4);Glob.	73 %	75 %

TABLE XV
EFFECTS OF GLOBAL POOLING AND EXTENDED CLASSIFIER

Global pooling	Added layers	Validation Accuracy	Test Accuracy
Global Max Pool	none	75 %	83 %
Global Max Pool	2D conv. (1,1) x64	75 %	81 %
Global Average Pool	none	73 %	75 %
Global Average Pool	2D conv. (1,1) x64	76 %	83 %

TABLE XVI
VARIOUS 3D CONVOLUTION MODELS

Model depth	Convolution kernels	Pooling kernels	Validation Accuracy	Test Accuracy
4	(9,3,3)stride of	(2,1,1);	60 %	57 %
+	(2,1,1);(9,3,3);	(2,1,1);		
2	(9,3,3);(12,3,3); best 2D model	(glob.,1,1); best 2D model		

D. Detachment defect optimized models using raw data

Experiments with raw data measurements did not reach good results. Reducing the input space by PCA slightly improved the results but still not good enough to be used. With only 30 true examples and higher difficulty in comparison to nails, it is very hard to design classification model for volumetric raw data. Due to PCA failing, we opted to design a new CNN model for reducing the input space from volumetric to planar data. This model is trained by using SCs images as ideal outputs. This way, there are much more data to lead training process correctly. The best model trained on SCs images are attached after this model. Our best results reached by combining these two modes are shown in TABLE XVI.

IV. CONCLUSION & FUTURE WORK

We presented AI models for automated classification of detachments and nails within paintings. We created small dataset containing 4 paintings measured with active thermography. Due to the small dataset size, we designed 4 specialized models optimized to the single defect type. The best nail classifier for SCs images reached 96.03 % accuracy and one for volumetric raw data measurements reached 93.65 % accuracy. The best detachment classifier for SCs images reached acceptable 87 % accuracy. However, the model optimized for volumetric raw data measurements reached only 57 % accuracy. The character of detachment, its more varied shapes and variations, its varied placement between different paint layers and varied texture character makes this a much more difficult task. Availability of only 30 true examples limited our options.

The presented models are designed for use before restoration process or for digitization purposes. They can reduce long manual work required for analysis of a large size work of Art. They also make the process of artwork analysis available for less experienced people in this domain.

Obtaining access to more paintings is a key part of our future work. Only with granted access, we can produce more

measurements and enlarge our dataset to improve detachment classification accuracy. We are hopeful and open to wider future research cooperation in this domain.

ACKNOWLEDGMENT

We highly appreciate and thank Barbara Balbi from the Superintendency of Archaeology, Fine Arts and Landscape for the city of Naples, and Maria Palma Recchia from the Superintendency for Archaeology, Fine Arts, and Landscape for the Metropolitan Area of Naples for kindly granting access to the paintings.

REFERENCES

- [1] Mishra, M. and Lourenço, P.B., 2024. Artificial intelligence-assisted visual inspection for cultural heritage: State-of-the-art review. *Journal of Cultural Heritage*, 66, pp.536-550.
- [2] Li, J., 2021, April. Application of artificial intelligence in cultural heritage protection. In *Journal of Physics: Conference Series* (Vol. 1881, No. 3, p. 032007). IOP Publishing.
- [3] Fontanella, F., Colace, F., Molinara, M., Di Freca, A. S., and Stanco, F. (2020). Pattern recognition and artificial intelligence techniques for cultural heritage. *Pattern Recognition Letters*, 138, 23-29.
- [4] Perera, Walpola Layantha; Messemer, Heike; Heinz, Matthias; Kretzschmar, Michael (2020): Detecting Treasures in Museums with Artificial Intelligence. Workshop Gemeinschaften in Neuen Medien (GeNeMe) 2020. Dresden: TUDpress
- [5] Z. Zhang et al., "Reducing Bias in AI-Based Analysis of Visual Artworks," in *IEEE BITS the Information Theory Magazine*, vol. 2, no. 1, pp. 36-48, 1 Oct. 2022, doi: 10.1109/MBITS.2022.3197102. keywords: Painting;Art;Image color analysis;Cultural differences;Feature extraction;Visualization;Imaging;Machine learning;Best practices;Systematics,
- [6] Girbacia, F. (2024). An Analysis of Research Trends for Using Artificial Intelligence in Cultural Heritage. *Electronics*, 13(18), 3738. <https://doi.org/10.3390/electronics13183738>
- [7] Angheluta, L. M., and Chiroasca, A. (2020). Physical degradation detection on artwork surface polychromies using deep learning models. *Rom. Rep. Phys.* 72(3), 805.
- [8] Mezina, A., Burget, R. and Kotrly, M. A deep learning approach for anomaly detection in X-ray images of paintings. *npj Herit. Sci.* 13, 127 (2025). <https://doi.org/10.1038/s40494-025-01724-9>
- [9] Rajic, N., Principal component thermography for flaw contrast enhancement and faw depth characterisation in composite structures, *Compos. Struct.* 58 (2002) 521–528.
- [10] Swita, R., Suszynski, Z. Processing of thermographic sequence using Principal Component Analysis, *Meas. Autom. Monit.* 61 (2015) 215–218
- [11] Rippa, M., Vigorito, M.R., Russo, M.R. et al. Active Thermography for Non-invasive Inspection of Wall Painting: Novel Approach Based on Thermal Recovery Maps. *J Nondestruct Eval* 42, 63 (2023). <https://doi.org/10.1007/s10921-023-00972-8>
- [12] Rippa, M., Pagliarulo, V., Napolitano, F., Valente, T., Russo, P. Infrared Imaging Analysis of Green Composite Materials during Inline Quasi-Static Flexural Test: Monitoring by Passive and Active Approaches, *Materials* 16, 3081 (2023).
- [13] C. Sun, A. Shrivastava, S. Singh and A. Gupta, "Revisiting Unreasonable Effectiveness of Data in Deep Learning Era," in 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 2017, pp. 843-852, doi: 10.1109/ICCV.2017.97. url: <https://doi.ieeecomputersociety.org/10.1109/ICCV.2017.97>
- [14] Hestness, J., Narang, S., Ardalani, N., Diamos, G.F., Jun, H., Kianinejad, H., Patwary, M.M., Yang, Y., and Zhou, Y. (2017). Deep Learning Scaling is Predictable, Empirically. *ArXiv*, abs/1712.00409.
- [15] Mumuni, A. and Mumuni, F. (2022). Data augmentation: a comprehensive survey of modern approaches. *Array*, 16, 100258. <https://doi.org/10.1016/j.array.2022.100258>
- [16] Maas, Andrew L., Hannun, Awni Y. and Ng, Andrew Y. Rectifier Nonlinearities Improve Neural Network Acoustic Models. In: *Proceedings of the 30th International Conference on Machine Learning (ICML)*. 2013, pp. 2–3. https://ai.stanford.edu/~amaas/papers/relu_hybrid_icml2013_final.pdf