Infrared thermal defect identification and reconstruction of artworks using a spatiotemporal deep neural network

M Moradi[1,2,*], R Ghorbani[3], S Sfarra[4], D M J Tax[3] and D Zarouchas[1,2]

[1] Structural Integrity & Composites Group, Aerospace Engineering Faculty, Delft University of Technology, Kluyverweg 1, 2629 HS Delft, The Netherlands
[2] Center of Excellence in Artificial Intelligence for structures, prognostics & health management, Aerospace Engineering Faculty, Delft University of Technology, Kluyverweg 1, Delft, 2629 HS, The Netherlands
[3] Pattern Recognition Laboratory, Faculty of Electrical Engineering, Mathematics and Computer Science, Delft University of Technology, Van Mourik Broekmanweg 6, 2628 XE Delft, The Netherlands
[4] Department of Industrial and Information Engineering and Economics (DIIIE), University of L'Aquila, Piazzale E. Pontieri 1, Monteluco di Roio, 67100 L'Aquila, AQ, Italy
[*] M.Moradi-1@tudelft.nl

ABSTRACT

Assessment of cultural heritage assets is now extremely important all around the world. The use of Infrared Thermography (IRT) is an interesting concept since surface and subsurface faults can be discovered by utilizing the 3D diffusion inside the object caused by external heat. The primary goal of this research is to detect defects in artworks, which is one of the most important tasks in the restoration of mural paintings. To this end, a spatiotemporal deep neural network is utilized for defect identification in a mock-up reproducing an artwork, taking into account both the temporal and spatial perspectives of step-heating thermography.

EXPERIMENTAL SETUP

The IRT-inspected artwork is a replica of Giotto's "Meeting at the Golden Gate" (a mural painting) that is preserved in Padua's Scrovegni Chapel (Italy).

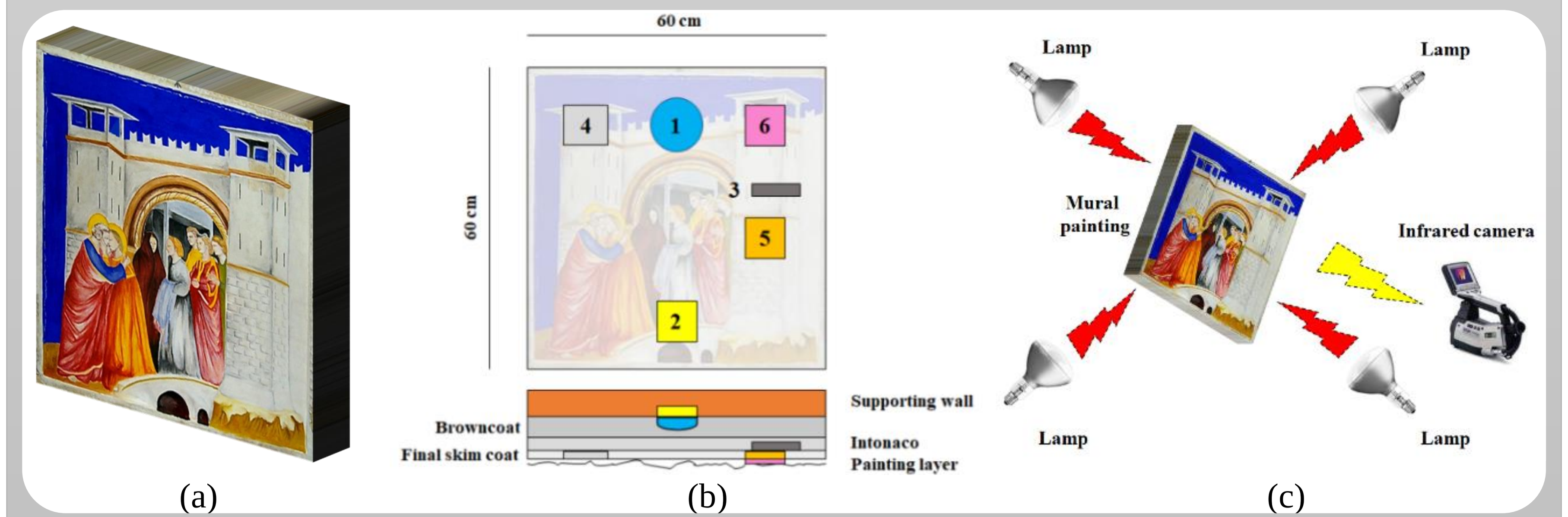

Figure 1. (a) Photograph of the mock-up, (b) defects map, and (c) sketch of the experimental IRT setup.

METHODLOGY

1) **Temporal network:** A multilayer perceptron (MLP) is used to classify the temporal signals related to the pixels into healthy or defective. The temporal sub-model comprises three hidden layers, respectively containing 20, 10, and 5 neurons with hyperbolic tangent as the activation function, whereas the logistic function is used in the output layer.
2) **Spatial network:** A U-Net is employed to segment images into healthy and faulty regions. The U-Net model's architecture is presented in detail in table 1.

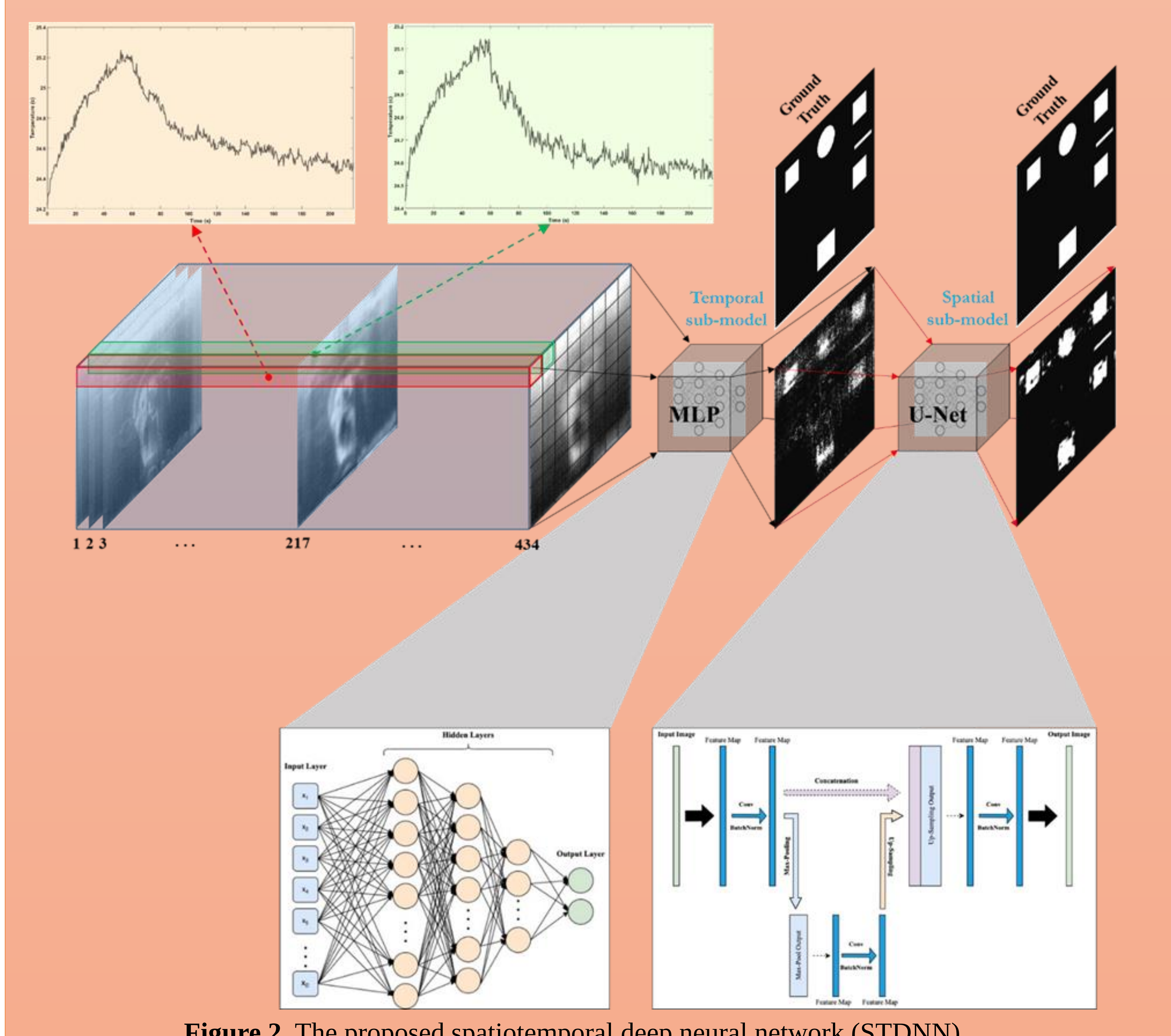

Figure 2. The proposed spatiotemporal deep neural network (STDNN).

Table 1. The U-Net (spatial sub-model) model's architecture.

Layer Number	Layer Type	Input of Layer
1	Conv2D (Filters: 128, (3 × 3), Activation: Exponential Linear Unit (ELU)), Kernel_Initializer: he_normal	MLP's output
2	Batch Normalization	Layer 1
3	Conv2D (Filters: 128, (3 × 3), Activation Function: ELU), Kernel_Initializer: he_normal	Layer 2
4	Max-Pooling (Size:(2 × 2))	Layer 3
5	Conv2D (Filters: 128, (3 × 3), Activation Function: ELU), Kernel_Initializer: he_normal	Layer 4
6	Batch Normalization	Layer 5
7	Conv2D (Filters: 128, (3 × 3), Activation Function: ELU), Kernel_Initializer: he_normal	Layer 6
8	Conv2D-Transpose (Filters: 128, (3 × 3), Strides = (2 × 2), Activation Function: ELU)	Layer 7
9	Conv2D (Filters: 128, (3 × 3), Activation Function: ELU), Kernel_Initializer: he_normal	Concatenation of Layers 3 and 8
10	Batch Normalization	Layer 9
11	Conv2D (Filters: 128, (3 × 3), Activation Function: ELU), Kernel_Initializer: he_normal	Layer 10
12	Dropout (0.3)	Layer 11
Final Layer	Conv2D (Filters: 1, (1 × 1), Activation: Sigmoid)	Layer 12

RESULTS AND DISCUSSION

The dataset, including 529 (23×23) patches, is randomly divided into training and test datasets with a ratio of 7:3 and repeated ten times to investigate the model's stability. As a result, the training and test datasets are composed of 370 and 159 patches, respectively, with each patch size of 10×10×434. Since this is a Class-Imbalanced problem, the AUC value is provided to analyze the performance of the sub-models. The AUC averaged over ten repeats across ten datasets of the MLP is presented in table 2. The results on the test set demonstrates the stability and high performance of MLP sub-model in classifying the pixels.

Table 2. AUC calculated from the (temporal sub-model) MLP's results for the ten generated datasets.

Dataset	1	2	3	4	5	6	7	8	9	10	Mean ± Pop std
Training	0.96±0.01	0.96±0.01	0.96±0.01	0.96±0.01	0.97±0.01	0.97±0.00	0.97±0.00	0.96±0.00	0.96±0.01	0.96±0.01	0.96±0.00
Test	0.86±0.01	0.89±0.01	0.87±0.01	0.88±0.01	0.88±0.00	0.87±0.01	0.89±0.01	0.87±0.01	0.89±0.01	0.86±0.01	0.88±0.01

The first repeat of the MLP's outputs from each dataset is selected as the input of U-Net model. To reduce the uncertainty in performance of U-Net model, this model is implemented ten different times on each of these datasets, and mean performances with their deviations are presented. Table 3 shows the U-Net model's performance across all ten datasets.

Table 3. AUC, Precision, Recall, and F1-Score calculated from the (spatial sub-model) U-Net's results for the ten generated datasets.

Dataset	AUC		Precision (Specificity)		Recall (Sensitivity)		F1-Score	
	Train	Test	Train	Test	Train	Test	Train	Test
1	1.00 ± 0.00	0.94 ± 0.00	0.97 ± 0.01	0.82 ± 0.03	0.96 ± 0.02	0.75 ± 0.03	0.96 ± 0.01	0.78 ± 0.01
2	1.00 ± 0.00	0.95 ± 0.01	0.91 ± 0.02	0.88 ± 0.02	0.97 ± 0.01	0.85 ± 0.02	0.93 ± 0.01	0.86 ± 0.01
3	1.00 ± 0.00	0.94 ± 0.00	0.95 ± 0.02	0.83 ± 0.04	0.94 ± 0.01	0.78 ± 0.03	0.95 ± 0.01	0.80 ± 0.01
4	0.99 ± 0.00	0.93 ± 0.01	0.92 ± 0.02	0.83 ± 0.01	0.95 ± 0.01	0.79 ± 0.03	0.93 ± 0.01	0.80 ± 0.02
5	0.99 ± 0.00	0.95 ± 0.00	0.95 ± 0.01	0.85 ± 0.02	0.94 ± 0.02	0.81 ± 0.02	0.94 ± 0.01	0.82 ± 0.01
6	1.00 ± 0.00	0.94 ± 0.00	0.89 ± 0.03	0.83 ± 0.02	0.97 ± 0.01	0.81 ± 0.02	0.93 ± 0.02	0.82 ± 0.01
7	1.00 ± 0.00	0.95 ± 0.00	0.96 ± 0.02	0.89 ± 0.03	0.96 ± 0.01	0.83 ± 0.03	0.96 ± 0.01	0.85 ± 0.01
8	0.99 ± 0.00	0.93 ± 0.00	0.92 ± 0.01	0.83 ± 0.02	0.90 ± 0.04	0.74 ± 0.05	0.91 ± 0.02	0.77 ± 0.03
9	1.00 ± 0.00	0.95 ± 0.00	0.96 ± 0.01	0.87 ± 0.02	0.94 ± 0.02	0.81 ± 0.01	0.95 ± 0.01	0.84 ± 0.01
10	1.00 ± 0.00	0.93 ± 0.00	0.92 ± 0.02	0.81 ± 0.02	0.95 ± 0.02	0.78 ± 0.03	0.94 ± 0.01	0.79 ± 0.01
Mean ± Pop std	1.00 ± 0.00	0.94 ± 0.01	0.94 ± 0.02	0.84 ± 0.03	0.95 ± 0.02	0.80 ± 0.03	0.94 ± 0.01	0.81 ± 0.03

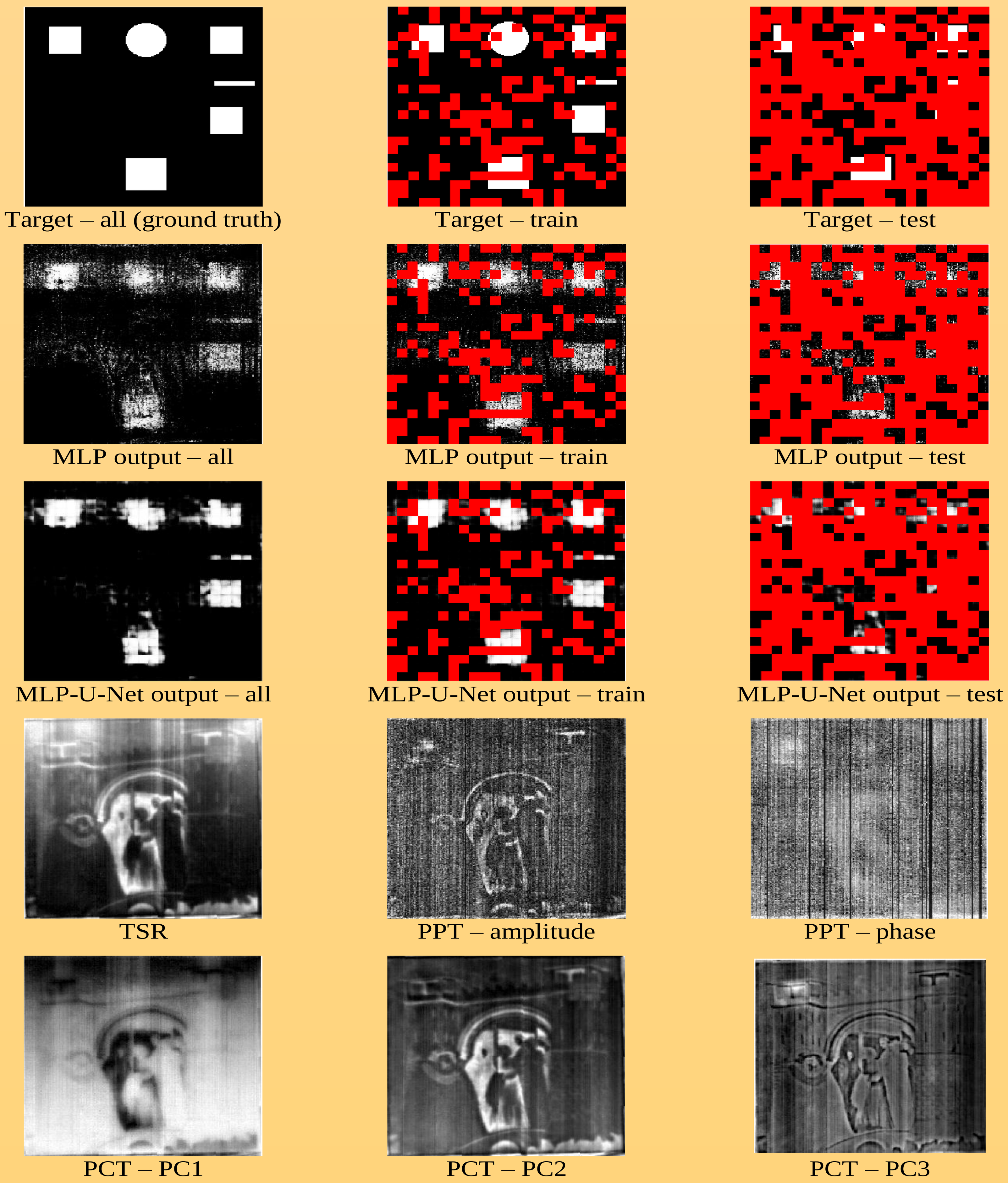

Figure 3. Results of MLP based on only temporal information (the 2nd row), MLP-U-Net based on temporal-spatial information (the 3rd row), the well-known algorithms of TSR, PPT, and PCT (the 4th and 5th rows) compared to the ground truth image (the 1st row) for the first dataset. The red color for the first three rows represents the unused pixels for training or test.

CONCLUSIONS

Initial results indicated that the mean F1-score evaluation metric is acceptable with a low standard deviation, which can be considered an outstanding performance despite the fact that there is a class-imbalance problem in the data. These results were supported by the AUC scores verifying that the model's performance was excellent and, more interestingly, stable. It was found that their results cannot be considered comparable to the MLP-U-Net's; for example, the effect and reflection of the drawing on the surface are still evident.