Paper No. 2391

Pattern Recognition Lab

2021 IEEE International Conference on Image Processing
19-22 September 2021 • Anchorage, Alaska, USA

GERMANISCHES NATIONAL MUSEUM

FAU FRIEDRICH-ALEXANDER UNIVERSITÄT ERLANGEN-NÜRNBERG
FACULTY OF ENGINEERING

CraquelureNet: Matching the Crack Structure in Historical Paintings for Multi-Modal Image Registration

Aline Sindel[1,2], Andreas Maier[1], and Vincent Christlein[1]

[1]Pattern Recognition Lab, FAU Erlangen-Nürnberg, Erlangen, Germany [2]Germanisches Nationalmuseum, Nuremberg, Germany

Contact: aline.sindel@fau.de https://lme.tf.fau.de/person/sindel/

Introduction

- Art investigations of paintings use different imaging systems: visual light photography (VIS), infrared reflectography (IRR), ultraviolet fluorescence photography (UV), and x-radiography (XR)
- Image registration to align the multi-modal images
- Visual features not necessarily visible by all modalities
- Use features of the crack structure due to their good visibility

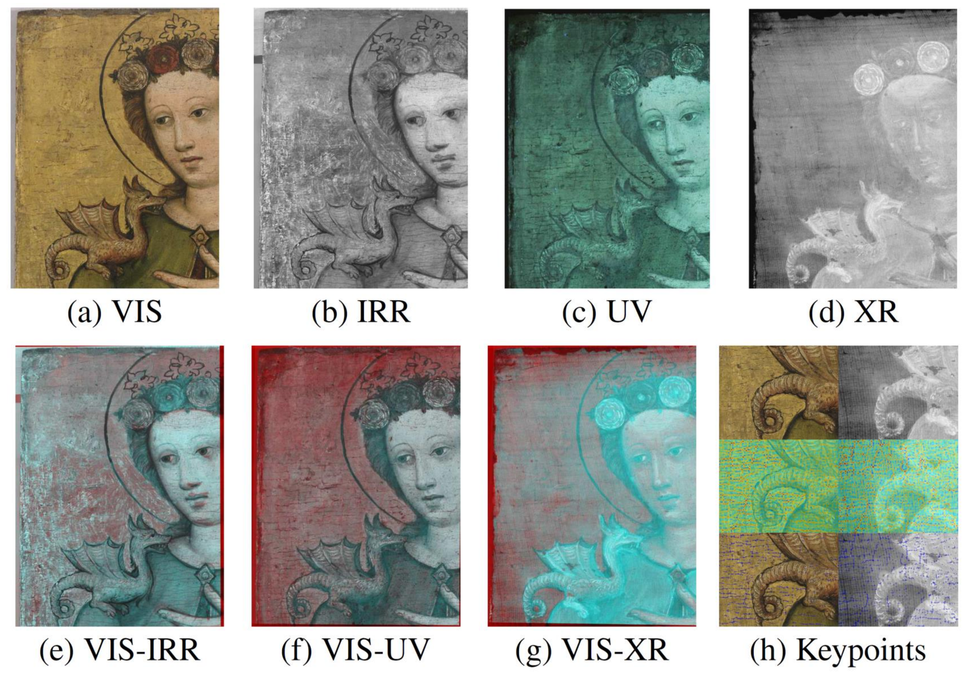


Figure 1: Multi-modal registration using CraquelureNet.

(a-d) Multi-modal images (VIS, IRR, UV, XR) of one painting

(e-g) Image fusion of registered images by CraquelureNet

(h) Heatmap of CraquelureNet and extracted keypoints for VIS-XR

CraquelureNet

CraquelureNet is a convolutional neural network (CNN) consisting of a ResNet [1] backbone and two heads.

Joint Training of Keypoint Detector and Descriptor Heads

$$\mathcal{L}_{\text{Total}} = \lambda_{\text{Det}} \mathcal{L}_{\text{CE}} + \lambda_{\text{Desc}} \mathcal{L}_{\text{QuadB}},$$

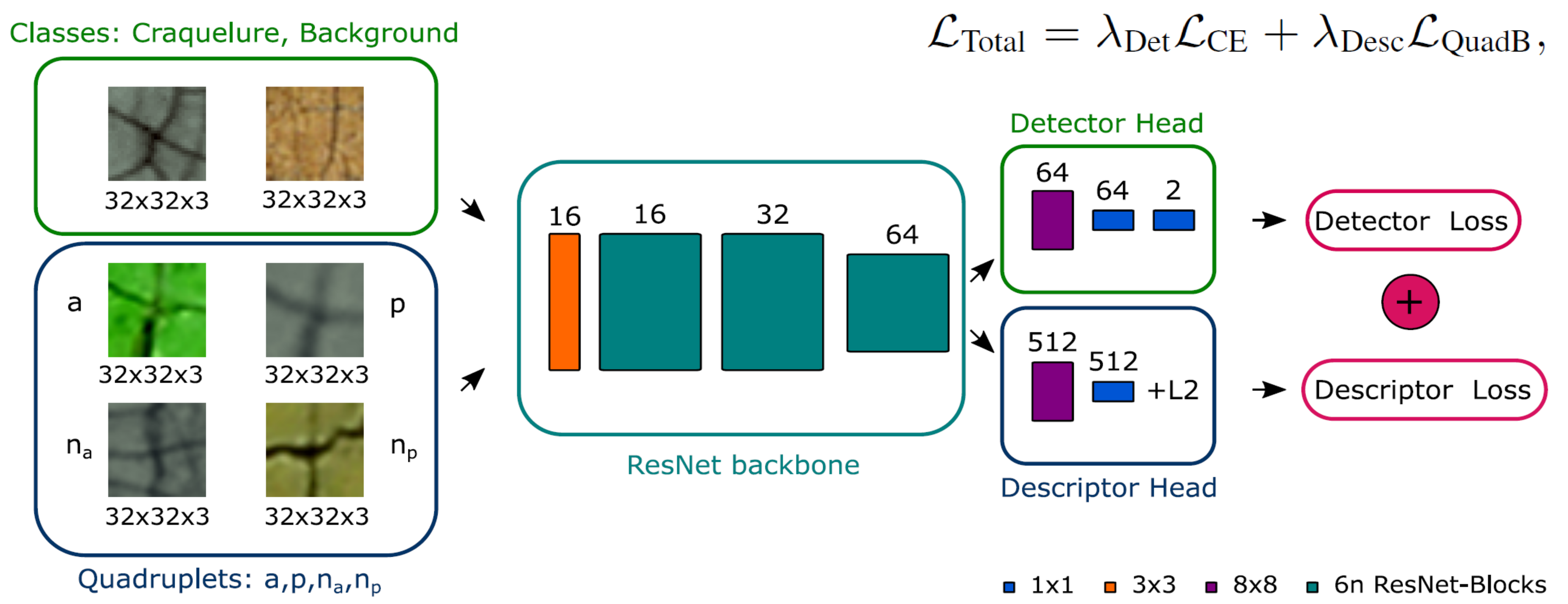


Figure 2: CraquelureNet: Joint training of detector and descriptor (patch-based).

- *Binary cross-entropy loss* for keypoint detector learning
- *Bidirectional quadruplet loss* for cross-modal descriptor learning

$$\mathcal{L}_{\text{QuadB}}(a,p,n_a,n_p) = \max[0, m + d(a,p) - d(a,n_a)] + \max[0, m + d(p,a) - d(p,n_p)],$$

CraquelureNet Inference

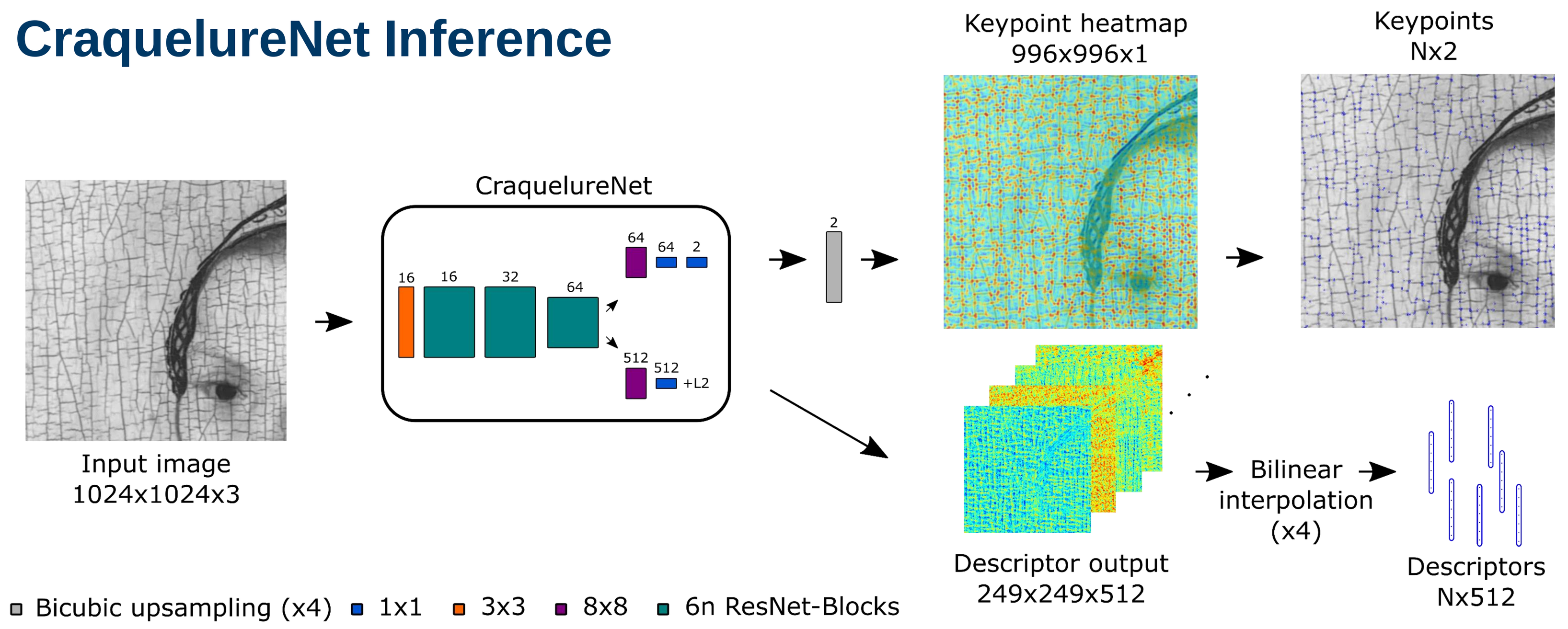


Figure 3: Inference of CraquelureNet using larger input sizes, extraction of keypoints and descriptors.

- Keypoints extracted from upscaled confidence heatmap
- Bilinear interpolation of descriptors at refined keypoint positions

Homography Estimation

- Mutual nearest neighbor matching of descriptors
- Estimation of homographies using RANSAC [2]

Results and Discussion

Multi-Modal Painting Dataset

- *Detection task:* 8730 (train) and 1992 (val) points for each modality (VIS, IRR, UV, XR) and class (craquelure, background)
- *Description task:* 5820 (train) and 2656 (val) point correspondences per domain (VIS-IRR, VIS-UV, VIS-XR)
- *Evaluation:* 15 (val) and 39 (test) image pairs with ground truth homographies computed using 40 point pairs each

Comparison of CraquelureNet to State of the Art

- Most correct matches for all domains (Fig. 4)
- Achieves highest success rate, repeatability, and matching score for all domains, with highest gain for VIS-XR (Tab. 1)

Table 1: Quantitative evaluation for the VIS-IRR, VIS-UV and VIS-XR test image pairs: Homography estimation (success rate of mean squared error of control points for error thresholds ε = {3, 5, 7}), detector repeatability (Rep), and RANSAC matching inlier score (MIR) at ε = 5.

Dataset	VIS-IRR					VIS-UV					VIS-XR				
Metrics [%]	Success rate			Rep	MIR	Success rate			Rep	MIR	Success rate			Rep	MIR
Error threshold	$\epsilon=3$	$\epsilon=5$	$\epsilon=7$	$\epsilon=5$	$\epsilon=5$	$\epsilon=3$	$\epsilon=5$	$\epsilon=7$	$\epsilon=5$	$\epsilon=5$	$\epsilon=3$	$\epsilon=5$	$\epsilon=7$	$\epsilon=5$	$\epsilon=5$
SIFT	23.1	23.1	23.1	18.5	6.5	15.4	23.1	23.1	21.8	12.9	0.0	0.0	0.0	14.9	0.8
D2-Net	15.4	53.8	**84.6**	19.6	37.4	23.1	46.2	61.5	20.4	36.3	0.0	15.4	15.4	14.6	11.0
AffNet+Hardnet	30.8	46.2	69.2	22.8	17.8	30.8	53.8	69.2	28.0	27.4	0.0	0.0	0.0	17.7	2.7
AffNet+Hardnet (fine-tuned)	30.8	61.5	61.5	23.2	20.2	30.8	38.5	38.5	28.2	28.3	0.0	0.0	0.0	17.6	3.4
SuperPoint	46.2	69.2	69.2	25.9	26.3	30.8	61.5	69.2	21.9	30.1	0.0	0.0	0.0	19.6	0.8
SuperPoint (fine-tuned)	30.8	38.5	38.5	22.9	17.1	30.8	38.5	46.2	18.8	17.9	0.0	7.7	7.7	18.9	1.1
CraquelureNet	**53.8**	69.2	**84.6**	43.5	**68.5**	**38.5**	**69.2**	**84.6**	**37.5**	**61.5**	23.1	38.5	53.8	**32.7**	**41.3**
CraquelureNet (in-domain)	**53.8**	**76.9**	**84.6**	**43.6**	63.5	23.1	**69.2**	**84.6**	36.8	58.7	**30.8**	**61.5**	**76.9**	31.9	37.0

Tested methods: SIFT [3], D2-Net [4], AffNet [5] + Hardnet [6], SuperPoint [7]; AffNet+Hardnet and SuperPoint were also fine-tuned on our multi-modal painting dataset

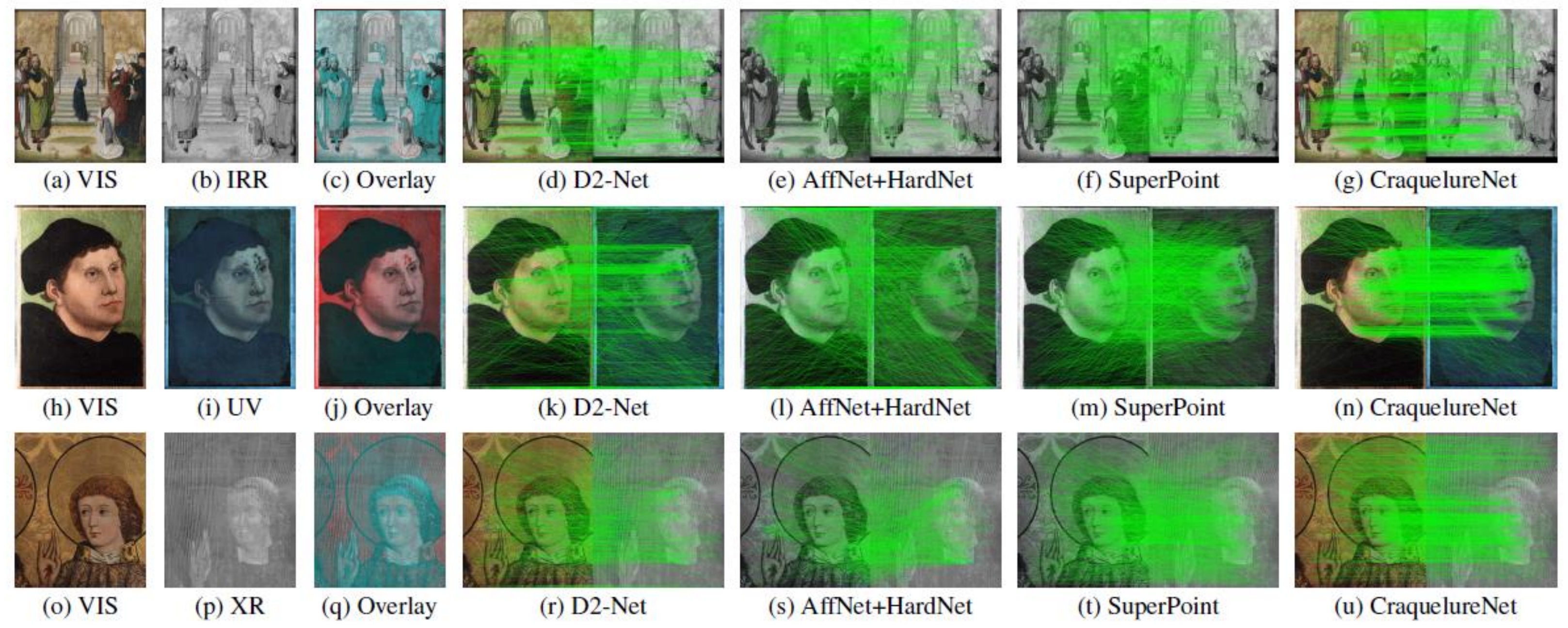
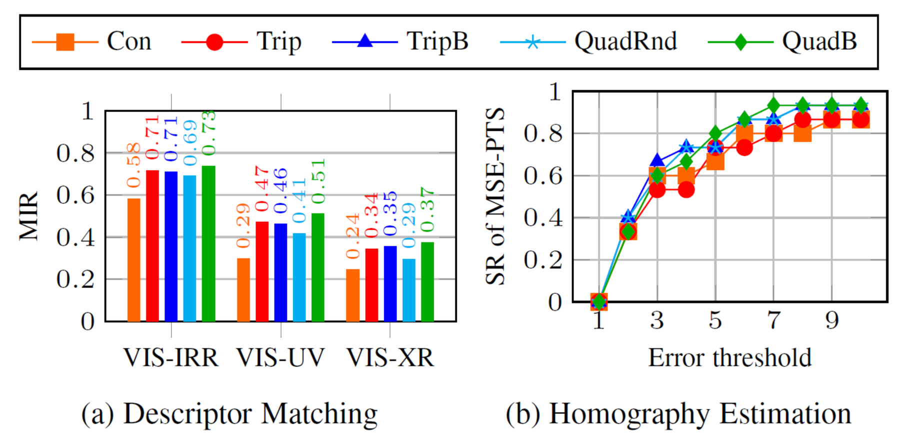


Figure 4: Qualitative results for image registration using CraquelureNet (c,j,q) and feature matching using CraquelureNet (g,n,u) or the competing methods for VIS-IRR, VIS-UV, and VIS-XR (test set)..

Influence of Descriptor Loss for CraquelureNet



Hard negative mining using bidirectional quadruplet loss:

- Highest RANSAC matching inlier ratio (MIR)
- Highest success rate (SR) of mean squared error of control points (MSE-PTS)

Figure 5: Influence of descriptor loss for the validation set: Contrastive loss (Con), Triplet loss (Trip), Bidirectional triplet loss (TripB) [6], Quadruplet loss with randomly selected fourth component (QuadRnd), and our bidirectional quadruplet loss (QuadB).

Conclusion

- CNN to jointly learn a cross-modal keypoint detector and descriptor using craquelure features
- Best registration results for the multi-modal dataset
- Future work: deep learning methods for the keypoint matching, outlier removal, and homography estimation

References

[1] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," Proc. IEEE CVPR 2016, pp. 770–778, 2016
[2] M. A. Fischler and R. C. Bolles, "Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography," Commun. ACM, pp. 381–395, 1981
[3] D. G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," IJCV, pp. 91–110, 2004
[4] M. Dusmanu, I. Rocco, T. Pajdla, M. Pollefeys, J. Sivic, A. Torii, and T. Sattler, "D2-Net: A Trainable CNN for Joint Description and Detection of Local Features," Proc. IEEE CVPR 2019, pp. 8084–8093, 2019
[5] D. Mishkin, F. Radenovic, and J. Matas, "Repeatability Is Not Enough: Learning Affine Regions via Discriminability," Proc. ECCV 2018, pp. 284–300, 2018
[6] A. Mishchuk, D. Mishkin, F. Radenovic, and J. Matas, "Working hard to know your neighbor's margins: Local descriptor learning loss," Proc. NIPS 2017, pp. 4826–4837, 2017
[7] D. DeTone, T. Malisiewicz, and A. Rabinovich, "Super-Point: Self-Supervised Interest Point Detection and Description," Proc. IEEE CVPR Workshops 2018, 2018