Puzzle Similarity: A Perceptually-Guided Cross-Reference Metric for Artifact Detection in 3D Scene Reconstructions

Project Page

Università della Svizzera italiana

idsia

Nicolai Hermann, Jorge Condor, Piotr Didyk

ICCV HONOLULU HAWAII OCT 19-23, 2025

Motivation

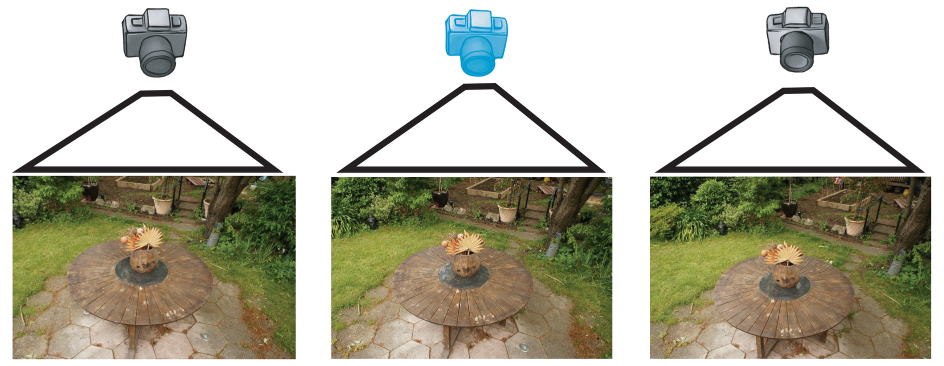
Current reconstruction techniques perform well when interpolating between training views once converged. However, scenes quickly fall apart when querying test views that do not fall onto the same training camera trajectory. Incongruent and unnatural artifacts become visible.

Scene Setup

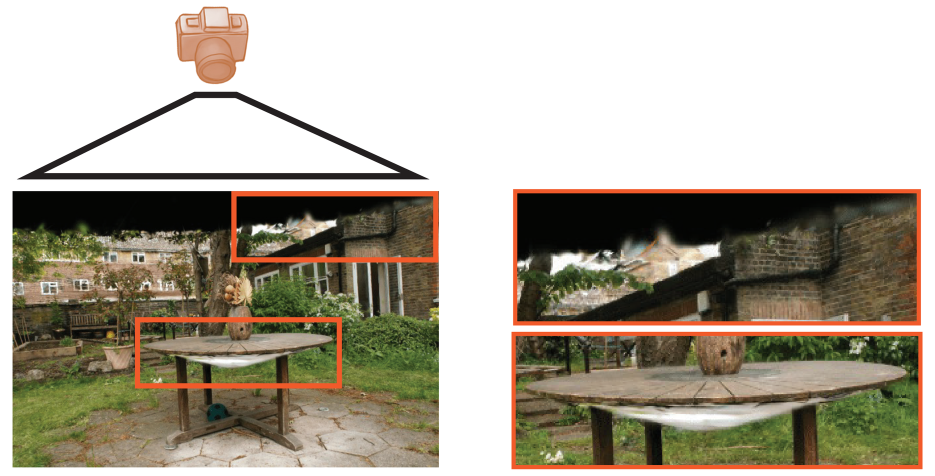
Training View

Validation View

Unseen View



Training view interpolation looks good!

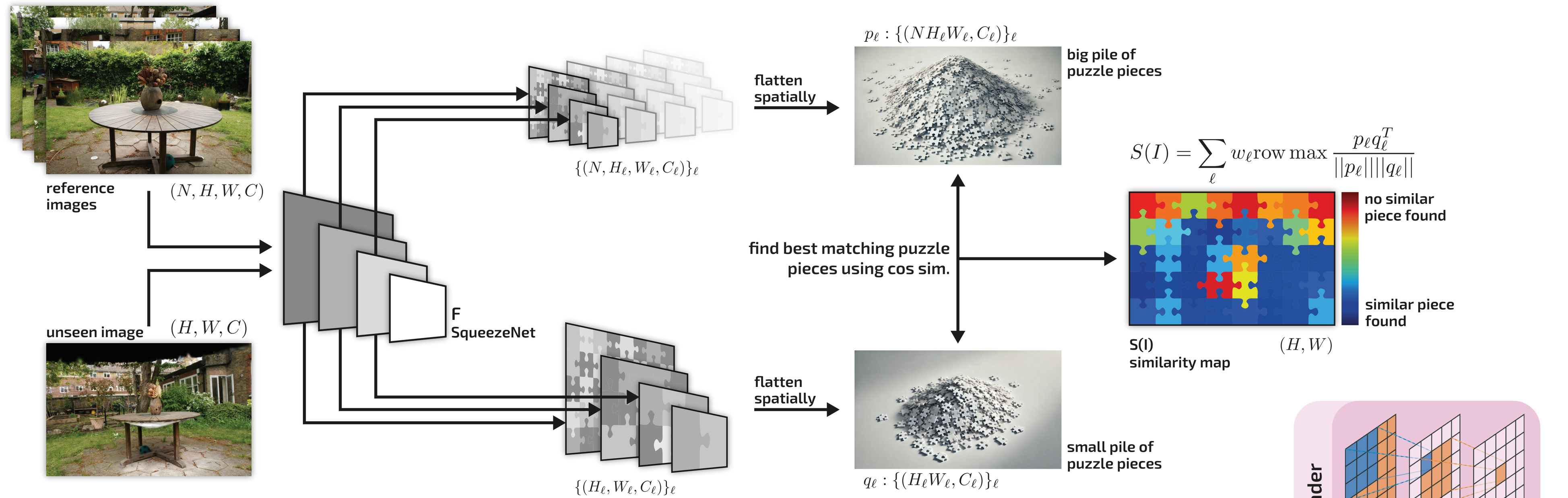


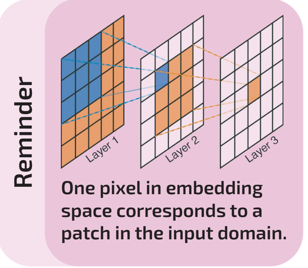
Serious artifacts in converged scenes

Method

```
$ pip install puzzle_sim
```

Following Zhang et al., we compare patches in embedding space (rather than image space) as this aligns better with human judgement. Thus, one puzzle piece corresponds to a single feature vector of some layer. For every layer ℓ, we compute the cosine similarity for every feature vector $q_{\ell,j}$ of the unseen image with all feature vectors p_ℓ from all training images. We can compute this in parallel using an outer product. The final similarity score for a pixel in the unseen image is given by the maximum similarity to any training feature.





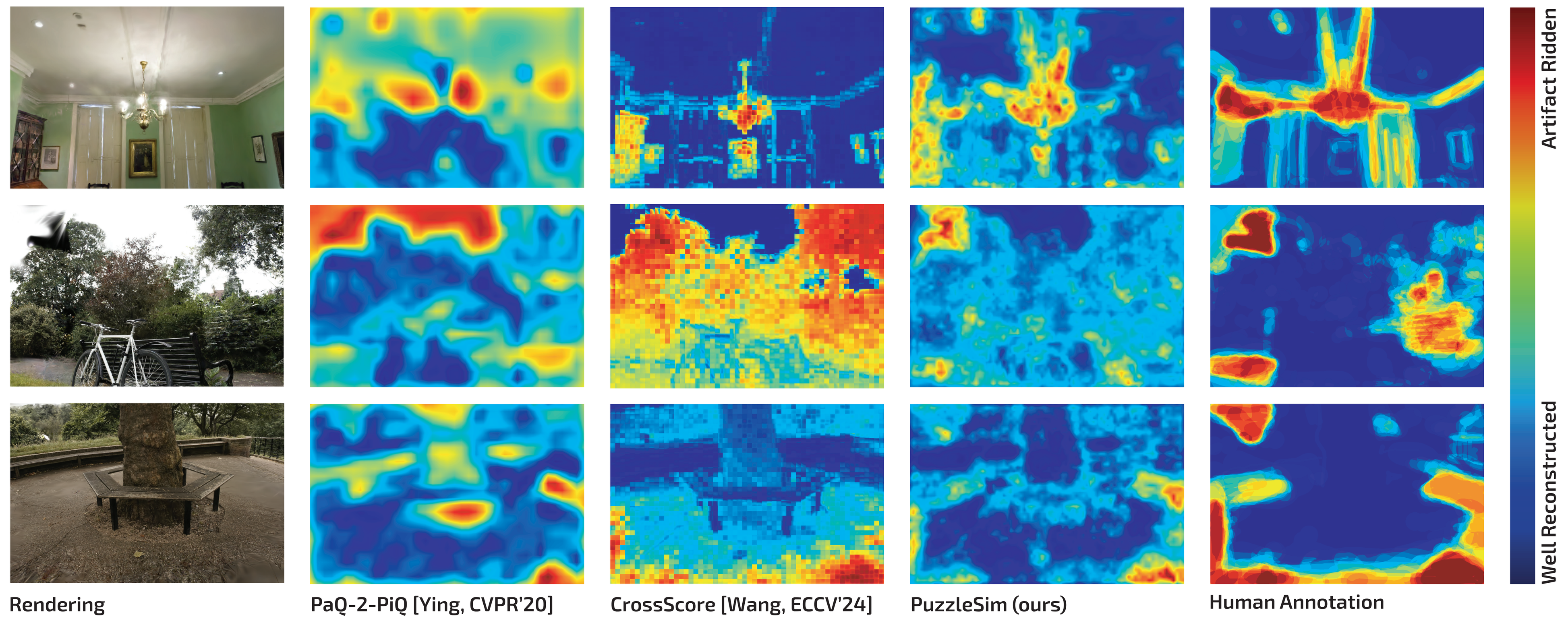
Every CNN backbone can be used to extract the embeddings. We also support the new Dinov3 backbones for an additional performance leap!

Validation

Metric	Pearson	Spearman
PAL4VST [Zhang, ICCV'23]	$0.078_{\pm 0.112}$	$0.062_{\pm 0.085}$
CNNIQA [Kang, CVPR'14]	$0.144_{\pm 0.247}$	$0.130_{\pm 0.253}$
PIQE [Venkatanath, NCC'15]	$0.292_{\pm 0.222}$	$0.268_{\pm 0.221}$
PaQ-2-PiQ [Ying, CVPR'20]	$0.402_{\pm 0.178}$	$0.349_{\pm 0.225}$
CrossScore [Wang, ECCV'24]	$0.510_{\pm 0.204}$	$0.378_{\pm 0.209}$
PuzzleSim (ours)	$0.615_{\pm 0.120}$	$0.474_{\pm 0.137}$

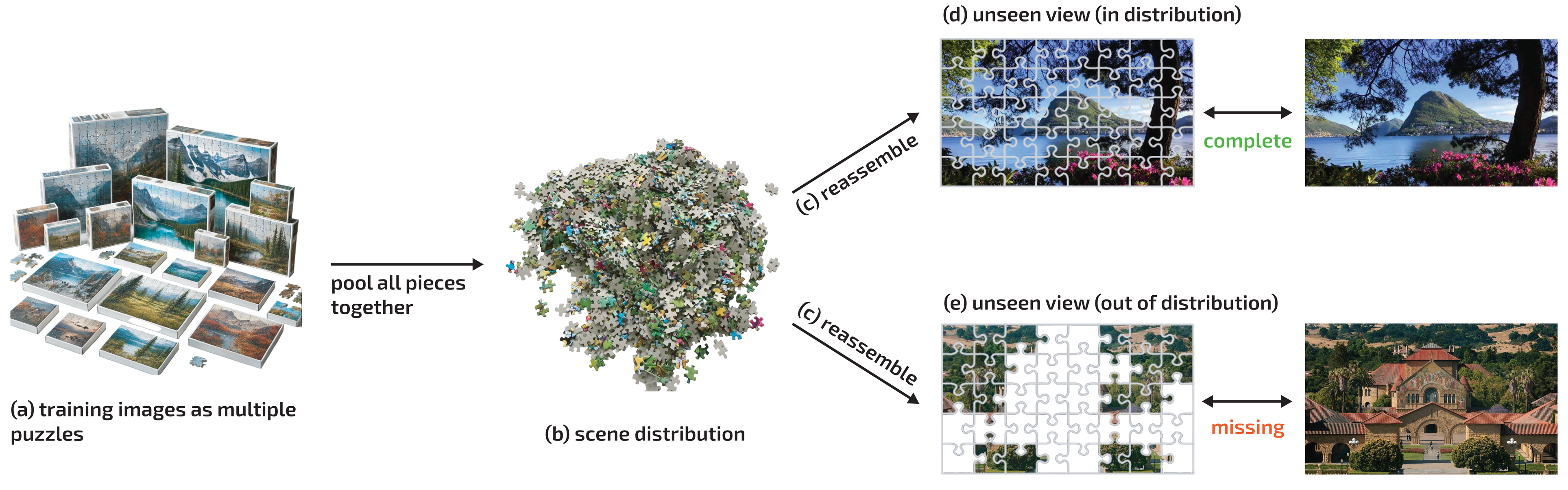
We compute pixelwise correlations between metric outputs and our collected human masks and find that our metric outperforms all competitors by a large margin.

Consequently, our error maps align much better with our human ground truth.



How to Find Artifacts in Unseen Views

There is no ground truth available to detect artifacts in unseen views so we leverage the training views (a). We extract patches from all training views and treat them as puzzle pieces (b). Using those puzzle pieces (patches) we try to reassemble an unseen view (c). There will be similar pieces for faithful reconstructions (d) but no similar pieces will be found for artifacts (e). This yields a partial reconstruction of the unseen view, where missing pieces mark artifacts.



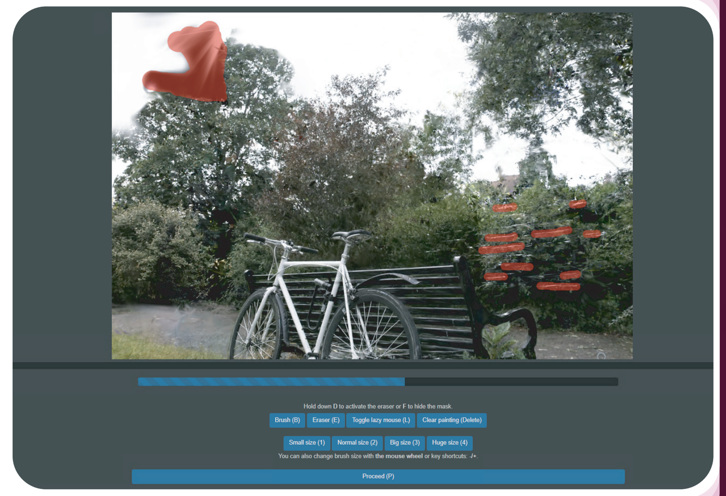
The lack of corresponding puzzle pieces suggests the presence of artifacts!

Source: visit.stanford.edu/assets/cardinal/images/explore-campus/architecture-landscape.jpg

Dataset

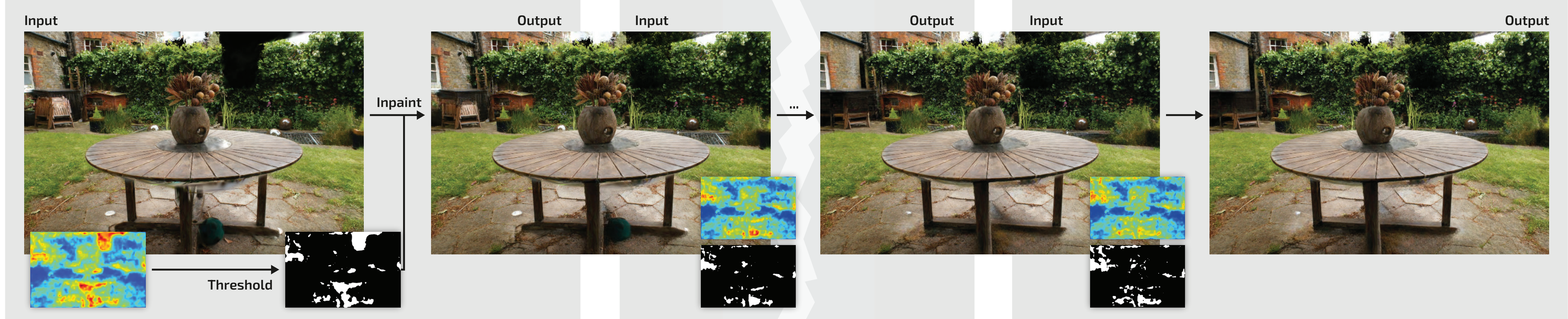
We collected a dataset of 36 artifact-ridden renderings over 12 challenging scenes, with varying artifact types and severities. We asked 22 participants to mark all areas that they found to be unnatural or unappealing, creating a binary mask (see the tool on the right). We estimated the probablitiy of observing an artifact as the average across all segmentations. The dataset is available on Hugging Face.

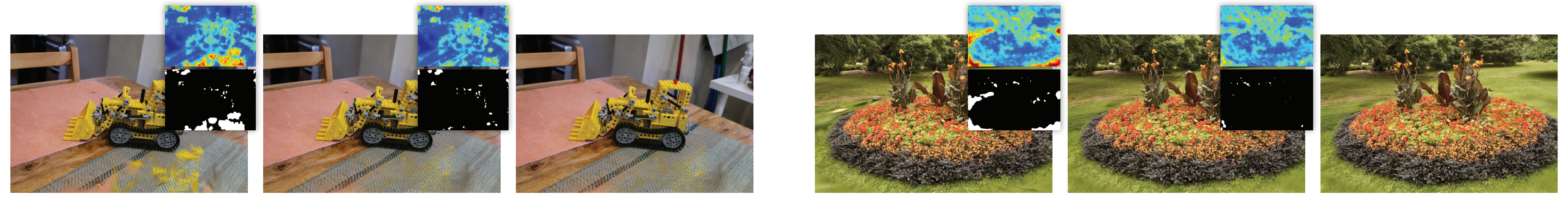
Get the data!



Automatic Inpainting

Artifacts and non-artifacts are linearly separable in our PuzzleSim maps. We propose an algorithm that thresholds the PuzzleSim maps, yielding outlier masks that we subsequently leverage, together with neural inpainting, to replace artifacts with more meaningful content. This process can be iterated until no further artifacts are detected.





Acknowledgements & Bibiliography

We would like to thank Krzysztof Wolski for making their image segmentation tool available to us, Volodymyr Kyrylov for providing the idea and first prototype of the memory-efficient implementation, and Sophie Kergaßner for design. This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement N°804226 PERDY), from the Swiss National Science Foundation (SNSF, Grant 200502) and an academic gift from Meta.

Richard Zhang, Phillip Isola, Alexei A. Efros, Eli Shechtman, and Oliver Wang. The Unreasonable Effectiveness of Deep Features as a Perceptual Metric.
Zirui Wang, Wenjing Bian, and Victor Adrian Prisacariu. CrossScore: Towards Multi-View Image Evaluation and Scoring.
Zhenqiang Ying, Haoran Niu, Praful Gupta, Dhruv Mahajan, Deepti Ghadiyaram, and Alan Bovik. From Patches to Pictures (PaQ-2-PiQ): Mapping the Perceptual Space of Picture Quality.
Lingzhi Zhang, Zhengjie Xu, Connelly Barnes, Yuqian Zhou, Qing Liu, He Zhang, Sohrab Amirghodsi, Zhe Lin, Eli Shechtman, and Jianbo Shi. Perceptual Artifacts Localization for Image Synthesis Tasks.
Le Kang, Peng Ye, Yi Li, and David Doermann. Convolutional Neural Networks for No-Reference Image Quality Assessment.
Venkatanath N, Praneeth D, Maruthi Chandrasekhar Bh, Sumohana S. Channappayya, and Swarup S. Medasani. Blind image quality evaluation using perception based features.