Reference-based Restoration of Digitized Analog Videotapes

Lorenzo Agnolucci[1], Leonardo Galteri[2], Marco Bertini[1], Alberto Del Bimbo[1]

[1]University of Florence – MICC
[1][name].[surname]@unifi.it

[2]Pegaso Telematic University
[2][name].[surname]@unipegaso.it

Media Integration and Communication Center

JAN 4-8 WACV 2024 WAIKOLOA HAWAII

Introduction

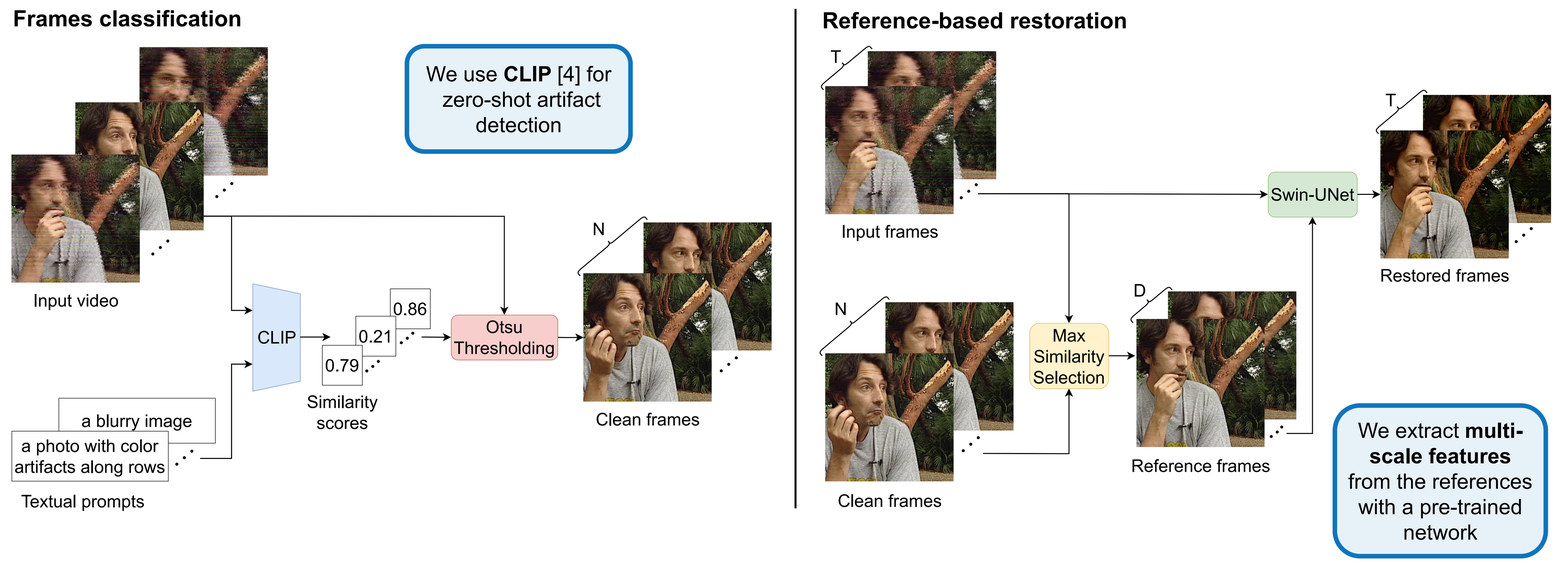
🔍 Analog magnetic tapes have been the main video data storage device for several decades, but their content shows unique and severe degradation

❌ Standard video restoration works are designed for digital videos and do not consider the artifacts caused by media issues, while old video restoration methods only focus on structured defects such as scratches

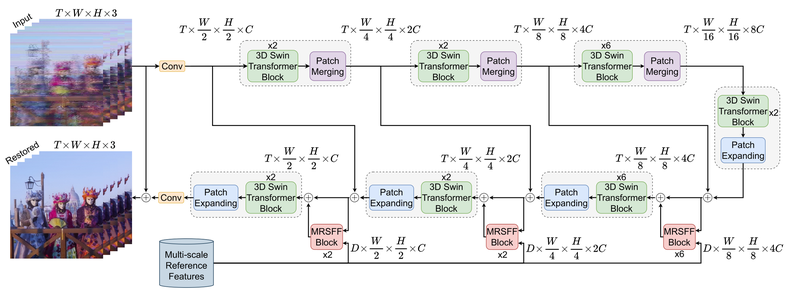
💡 We propose **TAPE**, an approach for restoring analog videos that exploits the time-varying nature of the artifacts by identifying the least damaged frames of each video with CLIP and employing them as references

💡 We develop a Swin-UNet architecture that leverages reference frames through our MRSFF blocks

TAPE Overview

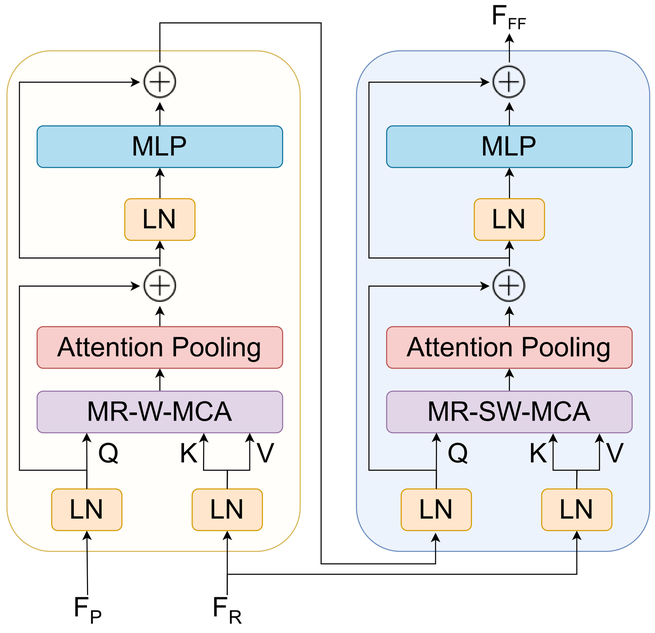


Architecture



- Our **Swin-UNet** architecture restores T frames at once by leveraging the information of both neighboring and reference frames
- The Swin Transformer [1, 2] allows taking advantage of the expressiveness of the attention mechanism while reducing the computational complexity
- Our Multi-Reference Spatial Feature Fusion (**MRSFF**) blocks rely on cross-attention and attention pooling to take advantage of the most useful parts of each reference frame
- Intuitively, each input frame looks at similar parts of the reference frames and exploits them to restore the lost details

MRSFF Block



Quantitative Results

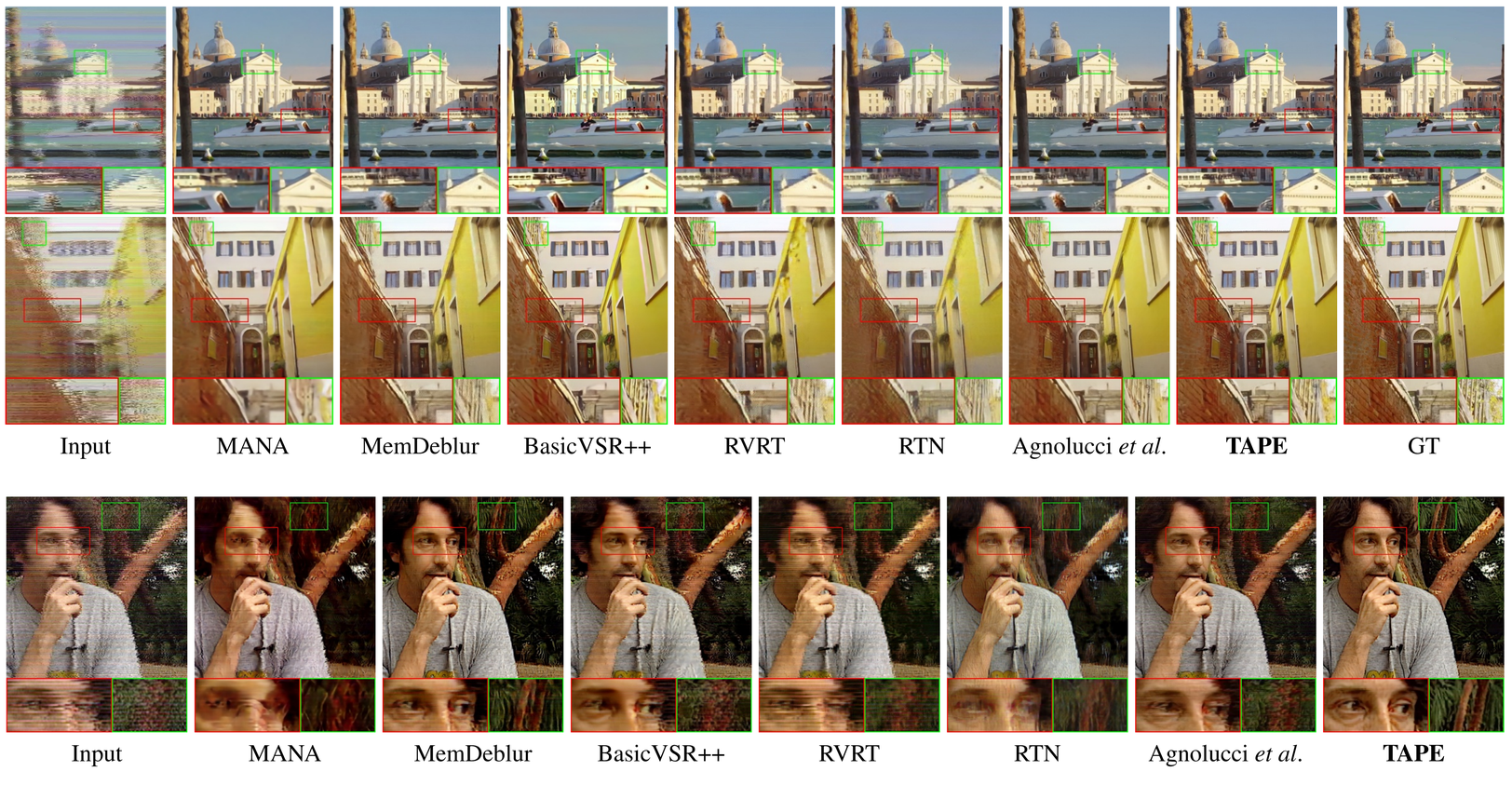
Synthetic dataset

Method	PSNR ↑	SSIM ↑	LPIPS ↓	VMAF ↑
MANA	27.81	0.843	0.206	40.28
MemDeblur	33.22	0.911	0.106	71.55
BasicVSR++	31.66	0.916	0.098	78.91
RVRT	32.47	0.896	0.117	72.41
RTN	31.46	0.905	0.100	56.76
Agnolucci *et al.* [3]	34.96	0.940	0.060	77.83
TAPE	**35.53**	**0.946**	**0.052**	**83.61**

Real-world dataset

Method	BRISQUE ↓	NIQE ↓	CONTRIQUE ↓
MANA	**41.80**	**5.90**	48.18
MemDeblur	51.20	8.89	45.82
BasicVSR++	59.19	8.42	48.44
RVRT	47.61	8.39	48.64
RTN	53.27	6.94	46.17
Agnolucci *et al.* [3]	59.44	7.90	45.45
TAPE	56.04	7.74	**42.99**

Qualitative Results



Conclusions

❌ Existing video restoration methods do not consider the artifacts typical of analog videos

💡 Our approach identifies the cleanest frames of each video using CLIP for zero-shot artifact detection and then exploits them as references through our MRSFF block

📈 TAPE achieves state-of-the-art results on both synthetic and real-world videos

📼 We release our synthetic dataset to foster research on this task

⚙ In future work, we will develop a learned degradation model to efficiently create more accurate synthetic videos

References

[1] Liu et al., "*Swin Transformer: Hierarchical Vision Transformer using Shifted Windows*", ICCV, 2021

[2] Liu et al., "*Video Swin Transformer*", CVPR, 2022

[3] Agnolucci et al., "*Restoration of Analog Videos Using Swin-UNet*", ACM MM, 2022

[4] Radford et al., *"Learning transferable visual models from natural language supervision*" ICML, 2021

arXiv

GitHub