Clément Maliet [1] Jérémy Boudret [1] Bouchra Harnoufi [2] Thierry Germa [1] Frédéric Claudel [2]

[1]Magellium

[2]Airbus Defence and Space

SUREDOS'24

30/05/2024

Assessment of Deep Learning Approaches for Satellite Video Super-Resolution

1, rue Ariane 31520 Ramonville Saint-Agne 05 62 24 70 00 info@magellium.fr https://www.magellium.com https://www.linkedin.com/company/magellium

- Satellite videos, although a promising market, suffer from a collection of quality and resolution related limitations which impede their adoption by the community.
- Advances in Video Super-Resolution (VSR) on natural imagery broke new grounds with the help of Deep Learning (DL) based approaches[1].
- On satellite VSR:
  - Few works could be found[2], notably due to the lack of readily available data.
  - Competitive results were reported but required significant modifications to adapt to the peculiar nature of satellite videos[3].
  - No common evaluation baseline exists → Metrics are computed on private datasets.

---

[1] H. Liu, Z. Ruan, P. Zhao, *et al.*, "Video Super Resolution Based on Deep Learning: A Comprehensive Survey," *Artif. Intell. Rev. 2022*, pp. 1–55, 2022. DOI: 10.1007/s10462-022-10147-y. arXiv: 2007.12928.

[2] Y. Luo, L. Zhou, S. Wang, *et al.*, "Video Satellite Imagery Super Resolution via Convolutional Neural Networks," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 12, pp. 2398–2402, 2017. DOI: 10.1109/LGRS.2017.2766204.

[3] Y. Xiao, X. Su, Q. Yuan, *et al.*, "Satellite Video Super-Resolution via Multiscale Deformable Convolution Alignment and Temporal Grouping Projection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2021. DOI: 10.1109/TGRS.2021.3107352.

Definition (Super Resolution problems)

$$\mathbf{Y} = \mathfrak{D}\left(\mathbf{X}; \theta_{\mathfrak{D}}\right) \downarrow_S \tag{1}$$

In practice we define SR as an ill-posed[4] inverse problem with regularization[5]:

$$\hat{\mathbf{X}} = \min_{\mathfrak{F}} \underbrace{\left\|\mathfrak{F}(\mathbf{Y}) - \mathbf{X}\right\|}_{\text{Data fidelity}} + \lambda \cdot \underbrace{\Psi\left(\mathfrak{F}(\mathbf{Y})\right)}_{\text{Prior}} \tag{2}$$

With $\mathfrak{D}$ often taken as a linear degradation plus a white noise[6].

[4] W. Yang, X. Zhang, Y. Tian, *et al.*, "Deep Learning for Single Image Super-Resolution: A Brief Review," *IEEE Trans. Multimed.*, vol. 21, no. 12, pp. 3106–3121, 2019. DOI: 10.1109/TMM.2019.2919431. arXiv: 1808.03344.

[5] S. Anwar, S. Khan, and N. Barnes, *A Deep Journey into Super-resolution: A Survey*, 2020. DOI: 10.1145/3390462. arXiv: 1904.07523.

[6] Z. Wang, J. Chen, and S. C. H. Hoi, "Deep Learning for Image Super-Resolution: A Survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 10, pp. 3365–3387, 2020. DOI: 10.1109/tpami.2020.2982166. arXiv: 1902.06068.

Videos extend (1) and (2), given a frame vector $\mathcal{N} = \{i - N, \dots, i + N\}$[7].

Definition (Video Super Resolution problems)

$$\mathbf{Y}_i = \mathfrak{D}\left(\mathbf{X}_i; \{\mathbf{X}_j\}_{j \neq i}^{j \in \mathcal{N}}; \theta_{\mathfrak{D}}\right) \downarrow_S \tag{3}$$

Let $u_{i \leftarrow j}$ the optical flow from j to i and $\mathbf{F}_{u_{i \leftarrow j}}$ the corresponding warp operator.

$$\hat{\mathbf{X}} = \min_{\mathfrak{F}} \operatorname*{argmin}_{\{\hat{u}_{i \leftarrow j}\}_{j \neq i}^{j \in \mathcal{N}}} \underbrace{\|\mathfrak{F}(\mathbf{Y}_i) - \mathbf{X}_i\| + \sum_{\substack{j \in \mathcal{N} \\ j \neq i}} \|\mathfrak{F}(\mathbf{Y}_j) - \mathbf{F}_{\hat{u}_{i \leftarrow j}} \mathbf{X}_j\|}_{\text{Data fidelity}} + \lambda \cdot \underbrace{\Psi\left(\mathfrak{F}(\mathbf{Y}_i)\right)}_{\text{Prior}} \tag{4}$$

[7] 1.

$$\hat{\mathbf{X}} = \min_{\mathfrak{F}} \operatorname*{argmin}_{\{\hat{u}_{i\leftarrow j}\}_{j\neq i}^{j\in\mathcal{N}}} \underbrace{\|\mathfrak{F}(\mathbf{Y}_i) - \mathbf{X}_i\| + \sum_{\substack{j\in\mathcal{N}\\ j\neq i}} \|\mathfrak{F}(\mathbf{Y}_j) - \mathbf{F}_{\hat{u}_{i\leftarrow j}}\mathbf{X}_j\|}_{\text{Data fidelity}} + \lambda \cdot \underbrace{\Psi\left(\mathfrak{F}(\mathbf{Y}_i)\right)}_{\text{Prior}} \tag{3}$$

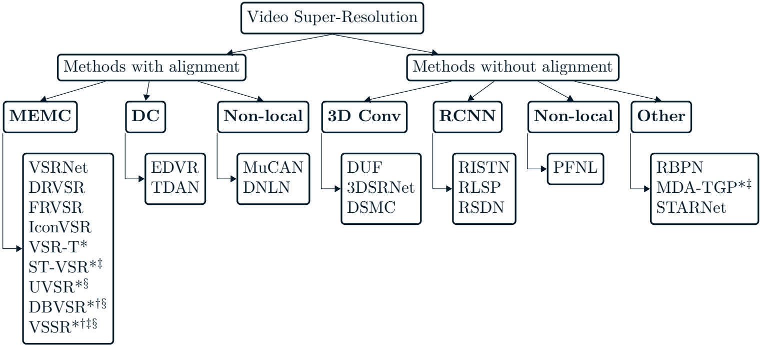
Corollary

With videos, understanding the interframe motion is key to getting good super-resolution performances and recover lost details.

Numerous works show that the bigger the neighborhood the better the reconstruction[7].

[7] K. C. K. Chan, S. Zhou, X. Xu, *et al.*, "Investigating Tradeoffs in Real-World Video Super-Resolution," *arXiv pre-print*, 2021. DOI: 10.48550/arxiv.2111.12704. arXiv: 2111.12704.

Adapted from the taxonomy in Liu et al.[8]:



[8] 1.

*Not reviewed by the Liu et al.

†Blind super resolution methods (i.e. not relying on knowing the degradation beforehand).

‡Indicates methods tailored for satellite videos.

§Indicates model-based methods as opposed to the more common learning-based architectures.

To alleviate data scarcity and to ensure we have a ground truth, we used simulated videos from aerial sequences provided by Airbus for the duration of the study.

Aerial videos

- 4K, 10Hz
- Stabilized over 5 seconds (50 frames) to avoid parallax-based deregistration effects
- 17cm or 24cm GSD

Simulations

- Supervised pairs (x4 SR):
  - At 30 cm (ground truth)
  - At 1m20 (observation)
- Online (during training) noise generation to improve generalization

Each sequence was spatially split in a training, validation, and testing set based on user-defined area-of-interest.

We used two complementary evaluation modalities:

- A quantitative assessment with common ”with reference” Image Quality Assessment (IQA) metrics (i.e. ℓ_1, ℓ_2, and SSIM):
  - The *de-facto* standard in the literature whenever a ground-truth is available.
  - Prone to texture deregistration[9] which unfairly penalizes high-frequency content[10].
  - Poorly correlated to human perception compared with dedicated metrics like SISR[11] or perceptually-trained ones like CORNIA[12].
- A qualitative assessment with a workshop organized with industrial partners.

---

[9] G. Jinjin, C. Haoming, C. Haoyu, *et al.*, “PIPAL: A Large-Scale Image Quality Assessment Dataset for Perceptual Image Restoration,” in *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 12356 LNCS, Springer Science and Business Media Deutschland GmbH, 2020, pp. 633–651. DOI: 10.1007/978-3-030-58621-8_37. arXiv: 2007.12142.

[10] M. Zhou, K. Yan, J. Pan, *et al.*, “Memory-augmented Deep Unfolding Network for Guided Image Super-resolution,” *arXiv pre-print*, 2022. DOI: 10.48550/arxiv.2203.04960. arXiv: 2203.04960.

[11] C. Ma, C. Y. Yang, X. Yang, *et al.*, “Learning a no-reference quality metric for single-image super-resolution,” *Comput. Vis. Image Underst.*, vol. 158, pp. 1–16, 2017. DOI: 10.1016/j.cviu.2016.12.009. arXiv: 1612.05890.

[12] P. Ye, J. Kumar, L. Kang, *et al.*, “Unsupervised feature learning framework for no-reference image quality assessment,” in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 2012, pp. 1098–1105. DOI: 10.1109/CVPR.2012.6247789.

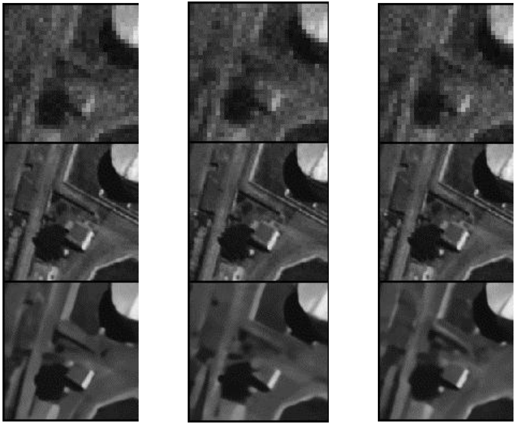
Test metrics

	ℓ_1	ℓ_2	SSIM
Classical†	29.27	3665	0.364
DNLN	**2.95**	**27.30**	**0.898**
EDVR	3.07	30.81	0.890
FRVSR	4.33	59.34	0.829
IconVSR	3.98	54.08	0.842
MDA-TGP	4.12	56.13	0.836
MuCAN	*3.03*	*29.23*	*0.892*
RSDN	3.49	38.44	0.867

Observation

Ground Truth

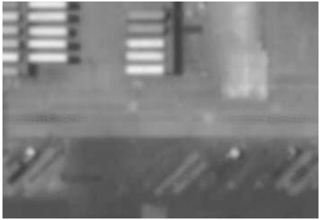
Predictions



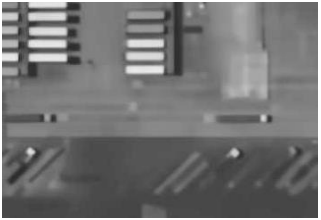
DNLN *MuCAN* *EDVR*

†A classical super-resolution method used as a baseline.

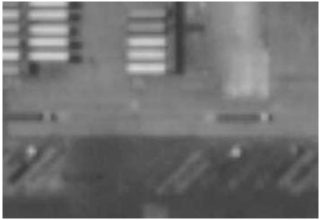
Results I



(a) Classical



(b) DNLN



(c) Observation

(d) Ground truth

DL VSR is able to take fine-grained motion into account like the moving trucks.

Results II

(a) Classical

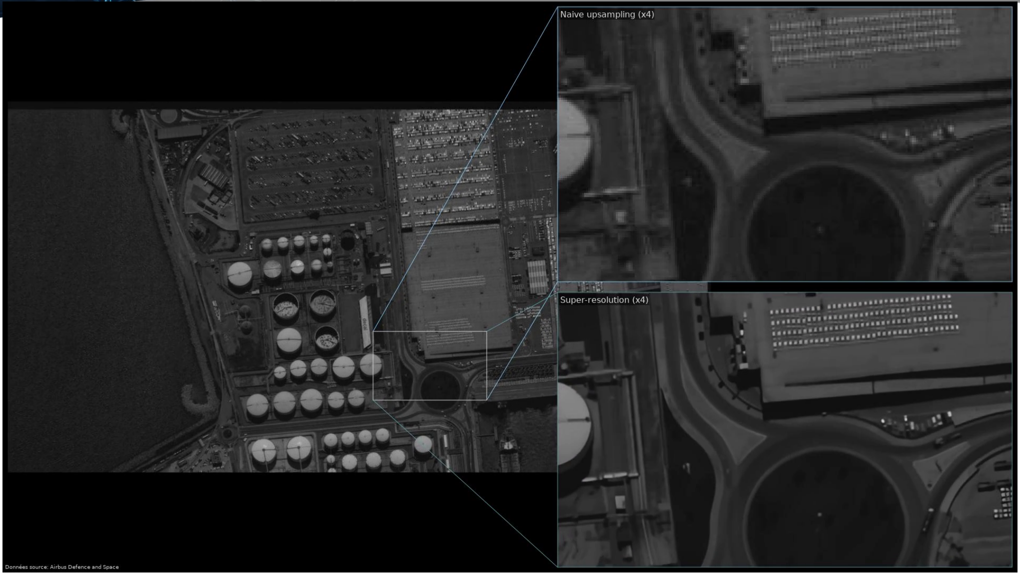
(b) DNLN

(c) Observation

(d) Ground truth

DL VSR avoids registration errors when some part of a sequence does not follow the global motion.

Results III


Naive upsampling (x4)
Super-resolution (x4)
Données source: Airbus Defence and Space

- This study showed that DL approaches for satellite video SR are indeed relevant and demonstrate competitive results:
  - Non-local methods emerged as clear winners in line with the de-blurring state-of-the-art.
  - Unintuitively, satellite-specific architectures under-performed compared with natural-oriented models.
- We demonstrated the technical feasibility of the method in a laboratory environment.
  - In particular, DL-based VSR methods showed significant improvement over classical algorithm.

Thierry Germa
Team Leader Space Robotics
thierry.germa@magellium.fr

Clément Maliet
Deep Learning Expert
clement.maliet@magellium.fr

(a) Classical

(b) DNLN

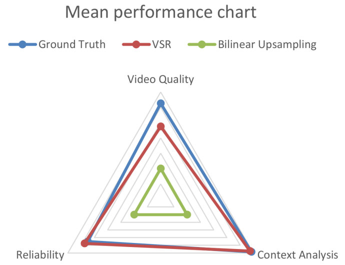
(c) Observation

(d) Ground truth

Classical VSR fails to reconstruct small moving cars.

Most common feedbacks

- VSR add legibility over simple interpolation.
- VSR could help reduce eyestrain for analysts.
- VSR improves the confidence in one's analysis.
- VSR demonstrated true lost details recovery in some cases.
- Small objects like pedestrians are irremediably lost.
- Recovered details felt consistent and robust.



Radar chart of the mean score obtained by sequence type in each category

[1] H. Liu, Z. Ruan, P. Zhao, *et al.*, "Video Super Resolution Based on Deep Learning: A Comprehensive Survey," *Artif. Intell. Rev. 2022*, pp. 1–55, 2022. DOI: 10.1007/s10462-022-10147-y. arXiv: 2007.12928.

[2] Y. Luo, L. Zhou, S. Wang, and Z. Wang, "Video Satellite Imagery Super Resolution via Convolutional Neural Networks," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 12, pp. 2398–2402, 2017. DOI: 10.1109/LGRS.2017.2766204.

[3] Y. Xiao, X. Su, Q. Yuan, D. Liu, H. Shen, and L. Zhang, "Satellite Video Super-Resolution via Multiscale Deformable Convolution Alignment and Temporal Grouping Projection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2021. DOI: 10.1109/TGRS.2021.3107352.

[4] W. Yang, X. Zhang, Y. Tian, W. Wang, J. H. Xue, and Q. Liao, "Deep Learning for Single Image Super-Resolution: A Brief Review," *IEEE Trans. Multimed.*, vol. 21, no. 12, pp. 3106–3121, 2019. DOI: 10.1109/TMM.2019.2919431. arXiv: 1808.03344.

[5] S. Anwar, S. Khan, and N. Barnes, *A Deep Journey into Super-resolution: A Survey*, 2020. DOI: 10.1145/3390462. arXiv: 1904.07523.

[6] Z. Wang, J. Chen, and S. C. H. Hoi, "Deep Learning for Image Super-Resolution: A Survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 10, pp. 3365–3387, 2020. DOI: 10.1109/tpami.2020.2982166. arXiv: 1902.06068.

[7] K. C. K. Chan, S. Zhou, X. Xu, and C. C. Loy, "Investigating Tradeoffs in Real-World Video Super-Resolution," *arXiv pre-print*, 2021. DOI: 10.48550/arxiv.2111.12704. arXiv: 2111.12704.

[8] G. Jinjin, C. Haoming, C. Haoyu, Y. Xiaoxing, J. S. Ren, and D. Chao, "PIPAL: A Large-Scale Image Quality Assessment Dataset for Perceptual Image Restoration," in *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 12356 LNCS, Springer Science and Business Media Deutschland GmbH, 2020, pp. 633–651. DOI: 10.1007/978-3-030-58621-8_37. arXiv: 2007.12142.

[9] M. Zhou, K. Yan, J. Pan, W. Ren, Q. Xie, and X. Cao, “Memory-augmented Deep Unfolding Network for Guided Image Super-resolution,” *arXiv pre-print*, 2022. DOI: 10.48550/arxiv.2203.04960. arXiv: 2203.04960.

[10] C. Ma, C. Y. Yang, X. Yang, and M. H. Yang, “Learning a no-reference quality metric for single-image super-resolution,” *Comput. Vis. Image Underst.*, vol. 158, pp. 1–16, 2017. DOI: 10.1016/j.cviu.2016.12.009. arXiv: 1612.05890.

[11] P. Ye, J. Kumar, L. Kang, and D. Doermann, “Unsupervised feature learning framework for no-reference image quality assessment,” in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 2012, pp. 1098–1105. DOI: 10.1109/CVPR.2012.6247789.