JOINT SUPER-RESOLUTION AND IMAGE RESTORATION FOR PLÉIADES NEO IMAGERY

F. Chouteau[1,*], L. Gabet[2], R. Fraisse[1], T. Bonfort[2], B. Harnoufi[1], V. Greiner[1], M. Le Goff[1], M. Ortner[1], V. Paveau[2]

[1] Airbus Defence and Space – Upstream Ground Systems, 31 Rue des Cosmonautes, 31400 Toulouse, France
(florient.f.chouteau, renaud.fraisse, bouchra.harnoufi, victor.greiner, matthieu.le-goff, mathias.ortner)@airbus.com
[2] Airbus Defence and Space – Geo Intelligence, 15 Rue des Satellites, 31030 Toulouse, France
(laurent.gabet, thomas.bonfort, vincent.v.paveau)@airbus.com

Commission I, WG I/1

KEY WORDS: Single-Image Super-Resolution, Pléiades Neo, Remote Sensing, Deep Learning, Convolutional Neural Networks

ABSTRACT:

Modern Earth Observation optical satellite systems, such as Airbus's Pleiades Neo (PNeo) push the boundaries of high spatial resolution by providing commercial imagery products with up to 30cm ground sampling distance (GSD). To further enhance the quality of the images, the in-space imaging system is usually complemented by on-ground image restoration processing, such as deconvolution and denoising (Latry et al., 2012). Recent advances leverage Convolutional Neural Networks (CNNs) to improve the image restoration quality (K. Zhang et al., 2021a).
Single Image Super-Resolution (SISR), or Zoom, the process of obtaining a higher resolution (HR) image from a lower resolved (LR) source, has recently gained traction for both medium resolution sensors such as Sentinel 2 (Lanaras et al., 2018) and high resolution such as Pléiades and GeoEye-1 (Zhu et al., 2020). This process further enhances the resolution of the image to improve downstream applications such as mapping (L. Zhang et al., 2021) and small objects recognition (Shermeyer and Van Etten, 2019). While SISR for remote sensing has been successfully tackled using CNNs (Rohith and Kumar, 2021) the main challenge for reaching acceptable image quality performance lies in the generation of realistic LR/HR training pairs (K. Zhang et al., 2021b).
In this paper, we propose:
- a dedicated simulation chain leveraging extremely-high-resolution (EHR) aerial imagery to generate realistic 30cm Pléiades Neo images and their corresponding fully restored HR equivalent at 15cm GSD
- A residual-based CNN architecture which we train to jointly restore and zoom the images

All contributions are assessed on real PNEO images. We deployed the trained models in a production context, to enhance the full Pléiades Neo products – with a swath of 47k pixels – in an efficient and scalable manner.

1. INTRODUCTION / CONTEXT

Satellite imagery is nowadays widely used and often a resource for analytics such as land use analysis, precision agriculture, natural disaster management, and other defence specific applications. These applications usually imply accurately detecting objects or characterizing specific land crops: spatial resolution is thus a key factor when deciding on which images to base the analytics on.

Therefore, being able to access better ground sample distance (GSD) for these analytics is critical. Hence, satellite imagery providers are competing to provide the best GSD. Unfortunately, enhancing resolution via imaging hardware improvement is expensive and technically challenging. The evolution of recent deep learning techniques toward even better performing algorithms is appealing and software-based image super-resolution (SR) techniques are now attractive in practice (Yang et al., 2019).

In recent years, the deep learning based super resolution domain has expanded and well performing super resolution models are now publicly available. But with those developments, their limits are now also more known. Some of the early models have shown limitations to specific deformations when embedded into mobile platforms. Moreover, generative model-based methods are to be closely monitored to prevent the inclusion of unwanted artifacts. In particular, contrary to mobile platforms, satellite platforms provide images with assumptions on their pixel content, and a naive super resolution approach would break these assumptions.

Target customers are really sensitive to these assumptions and a super resolution algorithm applied on satellite images should be designed adequately. In fact, super resolution models trained on natural images such as DIV2K tend not to generalize well on satellite images and require specific retraining on satellite images. There are a few attempts to train super resolution networks using satellite imagery (Rohith and Kumar, 2021). However, as stated in (Zhu et al., 2020), SISR approaches often uses a naive degradation model not representative of the true nature of remote sensing imagery. First, the noise model is of high importance and should be designed with care to represent real noise: the noise distribution of images acquired by satellite sensors vastly differs from white gaussian noise. Secondly, the Point Spread Function (PSF) in the satellite imaging system needs to be considered. Classical interpolation kernels are far from a good approximation of the real PSFs. Moreover, PSFs of most pushbroom sensors on satellites are spatially variant. Finally, variable motion blur caused by satellite movement or sensor scanning must be considered.

In this paper, we propose:
- an architecture on how to deploy a super resolution algorithm in an existing satellite processing chain,
- a realistic training data generation model leveraging extremely-high-resolution (EHR) aerial imagery to generate realistic 30cm Pléiades Neo images and their corresponding fully restored HR equivalent at 15cm GSD using proprietary measured parameters,

---

* Corresponding author

- a satellite-specific data augmentation,
- a Convolutional Neural Network (CNN)-based super-resolution model which can handle variant PSFs and different degrees of aliasing.

2. RELATED WORK

Image transformation tasks where a system receives some input image and transforms it into an output image are common in the deep learning landscape. Examples from image processing include denoising, super-resolution, and colorization, where the input is a degraded image (noisy, low-resolution, or grayscale) and the output is a high-quality image. The naive approach for solving image transformation tasks is to consider the problem as a supervised regression problem where a convolutional neural network is trained using a per-pixel loss function like L1 or L2 (Charbonnier et al., 1997). Such approaches have many advantages including their simplicity, their efficiency at test-time since they only require a forward pass through the trained network.

Recent developments have shown that per-pixel losses used by these methods do not capture perceptual differences between output and ground-truth images and loss functions should be adjusted to integrate overall image structure prior. In fact, classical loss functions do not represent the human perceived distance between images (Johnson et al., 2016). Perceptual loss functions have begun to appear: they are based not on differences between pixel values but instead on differences between high-level image feature representations extracted from pretrained convolutional neural networks (Johnson et al., 2016). In practice, they have played a major role in recent advances in super-resolution. At the beginning, image feature representations were extracted from pretrained networks on ImageNet and there is now an effort to design and train specific networks to be used as perceptual losses such as LPIPS (R. Zhang et al., 2018) or DISTS (Ding et al., 2020).

Regarding neural network architectures, residual-based architectures without pooling operations, pioneered by SRResnet (Ledig et al., 2017), and since then improved with more recent papers introducing complex blocks) are used in conjunction with UNets (K. Zhang et al., 2021a). More recently, Visual Transformers architectures have been introduced to the SISR problem (Liang et al., 2021). An important point to note is the positioning of the upsampling operation which differs between well-established architectures: the upsampling (usually a bilinear upsampling) step can occur before the network in pixel space, as in Zhu et al. (2020), whereas Ledig et al. (2017) uses subpixel upsampling after the residual network.

The data used for training are of extreme importance: first super resolution results worked on synthetic images but did not generalize. Nowadays, several datasets are publicly available such as DIV2K (Agustsson and Timofte, 2017) helping the super resolution community benchmark new solutions. Another key element to consider is the degradation model i.e., how to obtain both the Low Resolution (LR) and High-Resolution (HR) tuple. While the naïve approach consisted in downsampling and adding white gaussian noise, recently more realistic degradation models were formulated ((K. Zhang et al., 2021b) and (Wang et al., 2021)) to be able to train SISR model on purely synthetic LR/HR pairs. Unfortunately, such degradation models do not apply to satellite images : For example, in order to assess pretrained SISR models on remote sensing imagery, White et al. (2021) designed a specific degradation models matching their systems' sensors. In order to train SISR models for satellite imagery, a realistic simulation chain as well as aerial source images are needed to produce correct LR/HR training pairs : drawing from Zhu et al. (2020) and from our own experience, the design of the simulation chain is paramount and access to the sensor numerical model is essential to be able to train SISR models from synthetic data. Furthermore, the final objective is to produce good quality enhanced products of PNEO imagery. Simulation data is used to train the network, but final performance will be evaluated on PNeo imagery. This is a special case of domain adaptation where simulated data is to train and ground truth data (HR) cannot be accessed in the target domain. Currently, our work focuses on finding the best simulation parameters, but different semi-supervised techniques could also be used (such as Cycada (Hoffman et al., 2018)) to leverage simulation data and real data during training.

It should be noted that we are in the case where LR/HR pairs are available also known as aligned super resolution in contrast to the case where only 2 datasets are available: one LR and one HR with non-corresponding images in the two sets. Generative techniques are then well indicated for this problem as they do not require to have a tuple of LR and HR images (Lugmayr et al., 2019). Furthermore, for classical photography, state-of-the-art approaches include a loss term based on a discriminator network (Wang et al., 2021) along with L1 and perceptual losses. Unfortunately, these techniques do not provide the same level of control and are less-suited to remote sensing where the objective is enhance the information already in the image in terms of sharpness and signal-to-noise, rather than recreating "realistic" high-frequency information.

3. METHOD

3.1 Data

As stated before, a special care has been dedicated to the simulation chain to carefully design every numerical parameter to fit real images. We use Airbus owned Extremely High Resolution (EHR, 3 to 7.5cm GSD) aerial imagery in order to generate realistic LR/HR pairs of patches. The high scale factor between the aerial GSD and target GSDs limits the impact of the source sensor parameters in the simulation process, even when generating HR targets.

Using the knowledge of the sensor's characteristics (Point Spread Function (PSF), noise, radiometry factors) we generate (X_{LR}, Y_{HR}) training pair from Y_{EHR} aerial image with the following simulation model (eq. 1)

$$X_{LR} = ds(H_{LR} * R_{PNEO}(Y_{EHR}), GSD_{LR}), \qquad (1)$$
$$Y_{HR} = ds(H_{HR} * R_{PNEO}(Y_{EHR}), GSD_{HR}),$$
$$X_{LR} = \tilde{n}_{LR}(X_{LR}) = X_{LR} + N(0, \sqrt{a + b * X_{LR}}),$$

Where

- GSD_{LR} et GSD_{HR} are respectively input and target ground sampling distance,
- $ds(x, gsd)$ is an operator which downsamples x to a given gsd,
- H_{LR} is the PNeo PSF kernel,
- H_{HR} is a sharp anti-aliasing PSF kernel,
- $\tilde{n}_{LR}$ is a noise distribution approximating shot noise and depending on the pixel intensity,
- R_{PNEO} is a set of radiometric transformations mapping the source sensor dynamic to the PNeo sensor dynamic range.

The different Ground Sampling Distances (GSD_{LR} and GSD_{HR}) are chosen to be 30cm and 15cm in coherence with the objective to perform a x2 upsampling on PNEO images. The Point Spread Functions were chosen according to satellite sensor experts and were based on real acquisition measurements. H_{LR} is chosen to

represent real PNEO images, hence is based on real measurements. H_{HR} is chosen to generate a sharp target image but with limited aliasing and is sharper than H_{LR}. Thus, our problem formulation consists in a joint denoising, zoom and deconvolution process (Figure 2).

We show in Figure 1 examples of co-registered simulation and Pléiades Neo samples, illustrating that our simulation model closely matches the real images.
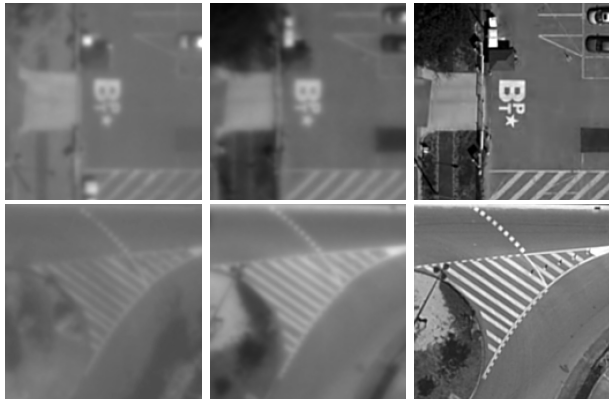

Figure 1 Image data examples. <u>Left</u>: Réal Pléiades Neo L1A, upsampled. <u>Center</u>: Simulated Pléiades Neo L1A, <u>upsampled</u>. <u>Right</u>: HR simulated target for super resolution
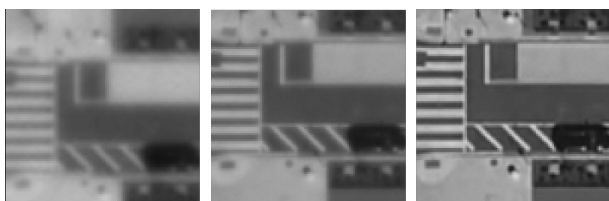

Figure 2 <u>Left</u>: Simulated LR image w/ noise, upsampled. <u>Centre</u>: Simulated HR image with the same PSF as the LR image, w/o noise. <u>Right</u>: Simulated HR target.

Note that the entire degradation model is far more complex than a bilinear downsampling and the addition of a gaussian distributed noise as classically done in super resolution on natural images. For instance, the noise function is a shot noise model where noise variation depends on the image content (Figure 3) but also depends on compression. This noise is added on-the-fly as data augmentation/domain randomization. During data augmentation, we also add specular reflection artifacts to increase variability and insert additional artifacts in the simulation process.

Aerial images were split to produce 2000x2000 LR images. A different blur level is selected in a predefined range to emulate uncertainty on the blur level of real images. In total, we were able to generate about 10k tuples of HR/LR patches of size 128 x 128 in LR resolution. In the future we plan to generate various Ground Sampling Distances and varying radiometric factors for each aerial image to increase our model's robustness to these kinds of uncertainties happening under certain acquisition conditions (e.g., off-nadir angle, illumination…).
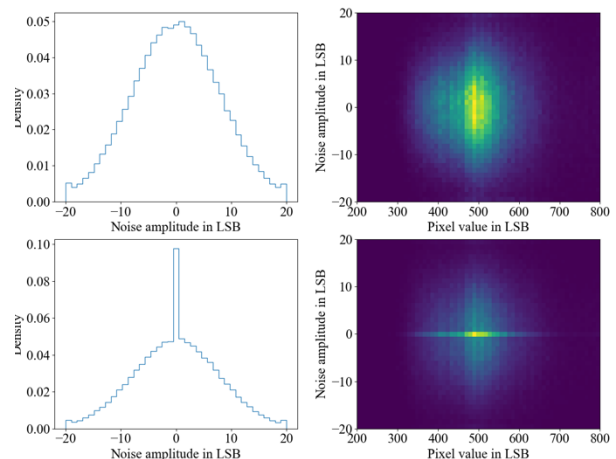

Figure 3 Noise distribution used data augmentation. First row is our data augmentation where the noise variance depends on the pixel LSB value. Second row is a white classical gaussian noise distribution. The problem is far more complex with our distribution and according to sensor models closer to reality

3.2 Problem Formulation

We choose to perform upsampling in pixel-space, as soon as possible in the processing chain and combine image restoration and zoom – effectively blind deconvolution since the network does not learn the upsampling operation – into a single CNN that replaces the nominal PNeo denoising and deconvolution (see (Latry et al., 2012) for a description of satellite image restoration applied to another sensor). We perform these operations in place of the nominal image restoration, in restored sensor geometry, before the orthorectification step (Figure 4). We also use the full sensor dynamic range (12bits). Such a choice leads to a simpler simulation model as we did not have to model the resampling operation, nor the image restoration process that we replace. The choice of architecture is based on several experiments and especially, performing the upsampling soon in the pipeline has shown promising results. Note that this implies that the denoising operation is done after having upsampled the noise distribution to the target GSD.
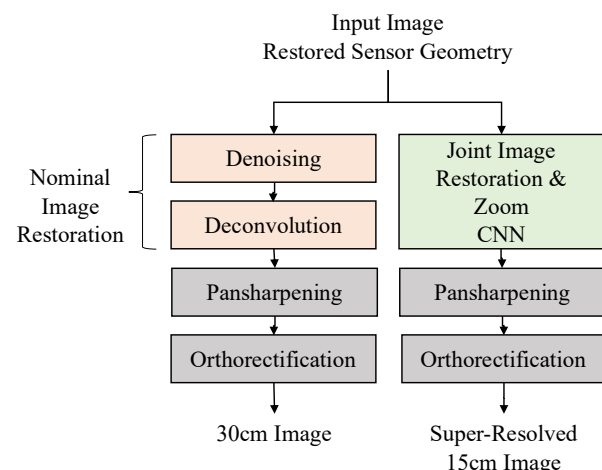

Figure 4 <u>Left</u>: Standard Image Processing Chain. <u>Right</u>: Proposed processing chain integration the joint image restoration and super-resolution network

Classically, a fully supervised learning approach is adopted where a network is trained to minimize a distance between the inferences from LR images and their ground truth HR. The network used is a residual-based CNN akin to SRResNet (Ledig

et al., 2017) with bilinear upsampling as a pre-processing step instead of late-stage pixel shuffle. We use a similar architecture than Zhu et al. (2020), with minor modifications such as 7x7 conv at the beginning and Channel Attention Block (Y. Zhang et al., 2018) in residual blocks. We use "replicate padding" instead of the classical "zero padding" in all convolutions to limit the border effect. See Figure 5 for the full network architecture.

As the network will be run in production on thousands of images, (a Pleiades Neo swath is around 47k pixels), the number of operations and weights were important to consider during the architecture choice, and that is why it was currently preferred over a classical and heavier Unet neural network. In practice, the inference takes place as denoted in (eq. 2):

$$\widehat{Y_{HR}} = f_\theta(h(X_{LR}))$$ (2)

Where:
- $\widehat{Y_{HR}}$ is the final high resolution image,
- X_{LR} is the low resolution input image,
- θ is the network parameters and f_θ is the final model,
- h is an upsampling operator which in our case is chosen to be a bilinear upsampling.
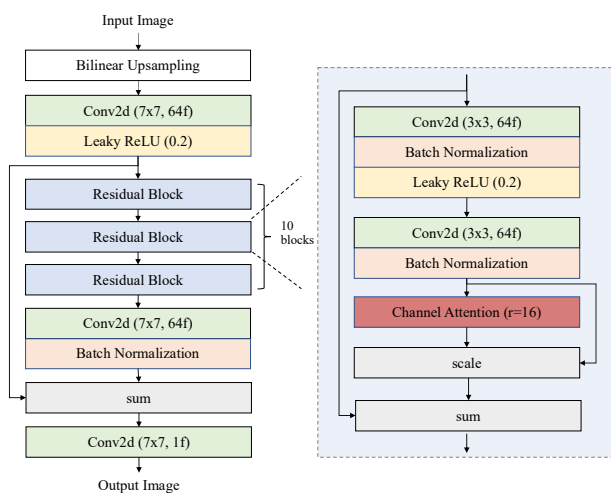

Figure 5 Our Neural Network Architecture.

After having defined the data and the model, we defined the loss to use in the optimization. The L1 loss function is often prescribed for super resolution problems but recent work shows that perceptual losses (Johnson et al., 2016) are quite effective for super resolution and strongly improve the visual aspect of inference images. For our trainings, we used the DISTS loss function (Ding et al., 2020), which combines the advantages of both the perceptual loss functions and the structural similarity index, as our perceptual loss component.

Our loss function is formulated as (eq. 3)

$$L(y_t, y_p) = \alpha\, SmoothL_1(y_t, y_p) + \beta\, DISTS(y_t, y_p)$$ (3)

where
- α and β are scalars tuned to the desired contribution for each loss term.
- $SmoothL_1$ is the smooth L1 loss function which is a combination of both the L1 loss for large deltas and the L2 loss for small deltas. The beta parameter of smooth L1 is set to 5 LSB.

We empirically set $\alpha = 0.5 * 1 / Mean(L1)_{10\ iterations}$ and γ identically so the contribution for both components is roughly equal after a few iterations.

For the sake of comparison, we also evaluated a loss based on the SSIM index which tries to improve image distances and close the gap between visual perception and mathematical image distance without using a pretrained network.

4. EXPERIMENTS

We used a dataset composed of simulation data based on the simulation process described in the paragraph above. The images used for the simulation are images of cities for most part but also some other landscapes (seas, etc.). We generated 7.6k training patches (before the data augmentation process) and 2.4k validation patches of size 128x128.

We evaluated different training strategies: different parameters for the SRRasnet architecture and the different loss functions described above (one parameter at each time). We trained the different SRResnets using the previously mentioned losses with a Lamb optimizer with a batch size of 16, an initial learning rate of 5.10^{-3} and a schedule strategy applying a 0.5 scale factor to the learning rate when the loss stagnates for more than 10 epochs. We ran the training for 200 epochs and kept the best validation checkpoint for each training. All trainings were run on an instance running on Google Cloud Platform with a single Nvidia-V100 GPU.

We kept two evaluation datasets aside from the training procedure:
- a dataset consisting of simulation data from other aerial images with simulation parameters similar but different from the ones used in training,
- a dataset consisting of hand selected PNeo images targeted for super resolution models evaluation.

On the evaluation dataset, where groundtruth is available, we compared the prediction and the ground truth using the multi-scale SSIM index, the PSNR, and the LPIPS index. The LPIPS index integrates a perceptual component and is used by the super-resolution community in complement to standard image processing metrics. For the dataset without groundtruth (real imagery) we used no-reference indices such as Total Variation and BRISQUE, since we do not have the ground truth available.

In practice, especially for super resolution, numerical metrics show strong weaknesses when compared to visual inspection e.g., a model can be visually pleasing while having poor metric values or the inverse when some artifacts are not reflected in the metric. As such, a strong emphasis is put on visual inspection to evaluate models and the last model selection step consists of validation by trained satellite image datasets. The evaluation is done by experts on the held out PNeo dataset where dedicated locations have been selected to compare super resolution models such as airports and ground patterns to evaluate sharpness and denoising properties of a given network.

5. RESULTS

5.1 Results

We report the results obtained by our method on simulated imagery (Figure 6). Images show very good results on both denoising and zoom operation, with fine details being restored without major artifacts. We also notice that some very fine details (which could only be seen at the native HR GSD) are not restored as they don't exist in the input image and thus are not "hallucinated" by the model.

We obtain similar results when applying our model to real images (Figure 7). This highlights the generalization power of a model trained on synthetic imagery. We find that the model outputs appear very stable under different acquisition conditions and landscapes, even those which were not present in the training dataset.
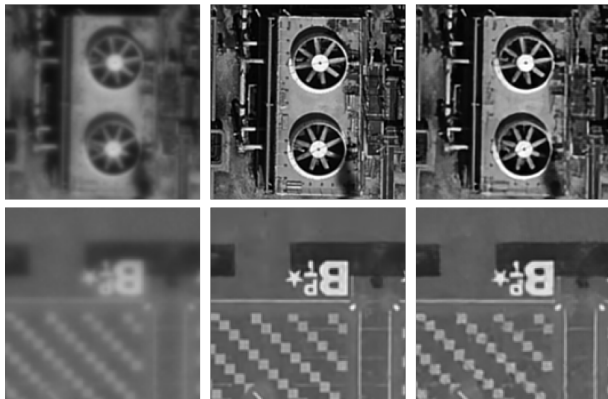

Figure 6 Results on simulated imagery. <u>Left</u>: Simulated LR image (30cm). <u>Centre</u>: Simulated HR image (15cm). <u>Right</u>: Output from model (15cm)
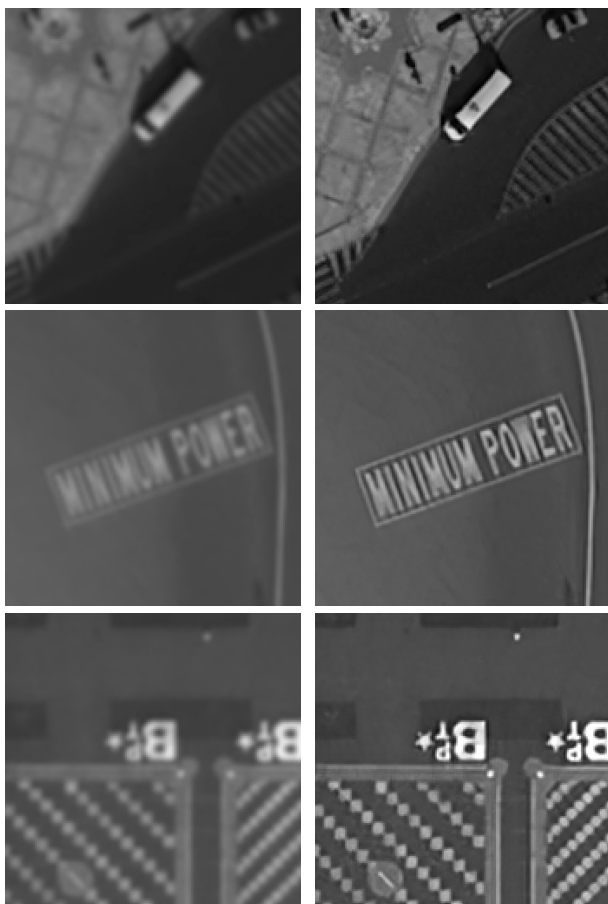

Figure 7 Results on PNeo imagery <u>Left:</u> Pleiades Neo L1B with nominal image restoration, after bicubic upsampling <u>Right:</u> Output from model prediction taking L1A (**unrestored** image) as input

We also report with-reference metrics on the simulated dataset (Table 1) as well as no-reference metrics on Pleiades Neo imagery (Table 2). For the full reference dataset, we compare our approach with a baseline of bicubic upsampling with a sharpening kernel. For the no-reference dataset, we compare our

approach to the nominal image restoration (denoising and deconvolution) of the Pleiades Neo images after a bicubic upsampling. Both metrics show a significant improvement over baselines.

	PSNR ↑	SSIM ↑	LPIPS↓
Bicubic upsampling + sharpening	36.48 dB	0.894	0.034
CNN (ours)	**42.65 dB**	**0.967**	**0.019**

Table 1 Image quality metrics comparing a bicubic upsampling followed by sharpening filter and our approach on the hold-out simulated dataset (64 x 64 patches)

	BRISQUE ↓	TV ↑
Nominal restoration + bicubic up.	82.65	0.92
CNN (ours)	**58.08**	**2.43**

Table 2 No-Reference image quality metrics on real Pléiades Neo images (64 x 64 patches) comparing the upsampled base restoration process and our CNN based approach. For BRISQUE, lower is better.

5.2 Ablation Studies

We ran ablation studies on the model architecture, lowering either the number of residual blocks (depth), the number of filters per block (width), and removing the squeeze-and-excite layer. We found that, while our baseline architecture remains the best according to the image quality metrics the differences remain small and computational complexity trade-offs may favour lighter models (Table 3). The perceived image quality remains, however, better for the more complex model, underlining the differences between the computational metrics and visual perception (Figure 8).

We also ran ablation studies on loss functions, using only L1 Loss, L1 and Multiscale SSIM and L1 and DISTS. We found that the results became blurrier and with more artifacts without the structural component and the perceptual loss (Figure 9).

Model	# params	throughput	PSNR ↑	SSIM↑
B=10, F=64, SE	788k	4.76s/Mpix	**42.65**	**0.967**
B=10, F=54, noSE	783k	3.95s/Mpix	41.90	0.962
B=5, F=64, SE	416k	2.25s/Mpix	42.36	0.965
B=10, F=32, SE	199k	2.26s/Mpix	42.23	0.966

Table 3 Ablation study on model architecture on the simulated imagery. We report throughput on a 1024x1024 tile split into 64x64 patches with 16 pixels of marging for the split-and-merge operation.
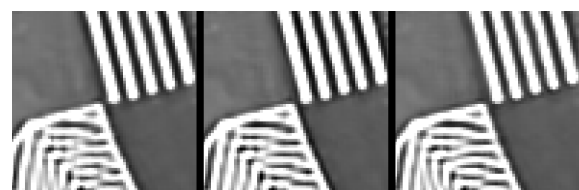

Figure 8 Architecture ablation results. <u>Left</u>: Shallower network. <u>Centre</u>: Narrower network <u>Right</u>: Main Network. Note the lines being better restored on the baseline model with less ringing effect and noise.

Figure 9 Loss function ablation results on real images. <u>Left</u> L1 Loss <u>Centre</u>: L1 + MS-SSIM <u>Right</u>: L1+DISTS. L1+DISTS appear sharper and with less artifacts and ringing effects.

5.3 Application to multispectral bands

While this article mainly focuses on the application of our approach to the Panchromatic band of Pleiades Neo, we adapted our method for the super-resolution of the PNeo multispectral channels, notably the blue, green, red, and near-infrared channels. The objective of such an approach is to be able to produce a better pansharpening by super resolving the MS channels before the fusion process, as MS channels at acquired at ¼ of the GSD of the panchromatic channels by the PNEo imaging system. We kept the same zoom factor (2) and applied a similar method with our dataset simulated at a GSD of 60/120cm, while keeping the 4 multispectral bands. We obtained similar results to the panchromatic approach with a similar network (Figure 10). This approach leads to the spectral components matching the panchromatic band more closely, thus making the colours appear sharper and without "smearing" effects.
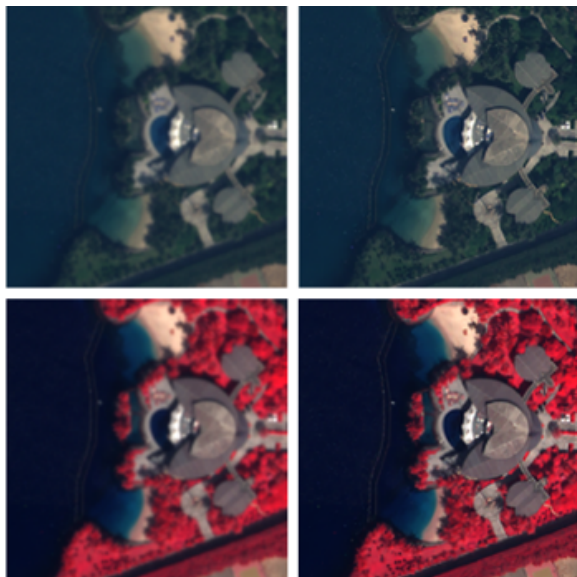

Figure 10 Results of our proposed approach on Pléiades Neo acquisition (multispectral bands at 120cm GSD). RGB and False Color Infrared are displayed. <u>Left</u>: L1B restored sensor with bicubic upsampling <u>Right</u>: proposed approach

5.4 Industrial Deployment

An end-to-end image processing chain capable of enhancing Pleiades Neo full strips at scale has been deployed in a public cloud environment. The processing chain is automatically activated once an image needing processing arrives on a dedicated bucket, by provisioning a docker container containing the described model on a GPU-enabled compute node, orchestrated by a Kubernetes cluster. The chain has been designed to run the classical image processing steps on CPU in parallel with the inference steps running on GPU, in order to maximize GPU usage rates. The super-resolved image is finally published to its destination bucket in cloud optimized geotiff format so that it is immediately available for online viewing.

6. CONCLUSION

In this paper we propose a realistic simulation process to generate a training dataset of paired LR/HR satellite images from very high-resolution aerial images. We show that this simulation process enables us to train a residual-based CNN architecture to jointly restore and zoom (single-image super resolution) its input image, and that it generalizes well to real satellite imagery. Thus,

our process enables the development of image restoration models from purely synthetic training datasets.

We highlight that image quality metrics on the simulated dataset may not reflect the final perceived quality and usability of the model on real imagery, a point to keep in mind when selecting the best model to deploy. It is likely that the final choice between several models will be done with a blind image quality assessment by qualified image quality engineers.

We applied this method to obtain a joint super-resolution and image restoration model that can be applied on Pleiades Neo imagery, and we developed a complete image processing chain for an operational deployment of the single-image super-resolution module in a production context, where the CNN is complemented by subsequent image production steps, such as orthorectification, pansharpening and local contrast adjustments, to produce analysis-ready 8-bit imagery.

In the future, we will focus on improving our simulation model to ensure that our model is robust to the full range of the images that can be imaged by the PNeo satellites. We will also explore a full radiometric image quality assessment (Yalçin et al., 2021) for the generated super-resolved products as well as improvements to the approach, such as implementing a more progressive image restoration pipeline by separating the denoising and the zoom steps with several networks chained together and learned sequentially – and designing a hard example mining procedure for the SR problem to be able to focus on more difficult examples during training as not all patches are equals for the SR problem.

REFERENCES

Agustsson, E., Timofte, R., 2017. NTIRE 2017 Challenge on Single Image Super-Resolution: Dataset and Study. Presented at the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 126–135.

Charbonnier, P., Blanc-Féraud, L., Aubert, G., Barlaud, M., 1997. Deterministic edge-preserving regularization in computed imaging. IEEE Transactions on image processing 6, 298–311.

Ding, K., Ma, K., Wang, S., Simoncelli, E.P., 2020. Image Quality Assessment: Unifying Structure and Texture Similarity. IEEE Trans. Pattern Anal. Mach. Intell. 1–1. https://doi.org/10.1109/TPAMI.2020.3045810

Hoffman, J., Tzeng, E., Park, T., Zhu, J.-Y., Isola, P., Saenko, K., Efros, A., Darrell, T., 2018. CyCADA: Cycle-Consistent Adversarial Domain Adaptation, in: Proceedings of the 35th International Conference on Machine Learning. Presented at the International Conference on Machine Learning, PMLR, pp. 1989–1998.

Johnson, J., Alahi, A., Fei-Fei, L., 2016. Perceptual Losses for Real-Time Style Transfer and Super-Resolution, in: Leibe, B., Matas, J., Sebe, N., Welling, M. (Eds.), Computer Vision – ECCV 2016. Springer International Publishing, Cham, pp. 694–711. https://doi.org/10.1007/978-3-319-46475-6_43

Lanaras, C., Bioucas-Dias, J., Galliani, S., Baltsavias, E., Schindler, K., 2018. Super-resolution of Sentinel-2 images: Learning a globally applicable deep neural network. ISPRS Journal of Photogrammetry and Remote Sensing 146, 305–319. https://doi.org/10.1016/j.isprsjprs.2018.09.018

Latry, C., Fourest, S., Thiebaut, C., 2012. Restoration Technique for Pleiades-Hr Panchromatic Images. ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences 39B1, 555–560. https://doi.org/10.5194/isprsarchives-XXXIX-B1-555-2012

Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A., Tejani, A., Totz, J., Wang, Z., 2017. Photo-realistic single image super-resolution using a generative adversarial network, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 4681–4690.

Liang, J., Cao, J., Sun, G., Zhang, K., Van Gool, L., Timofte, R., 2021. SwinIR: Image Restoration Using Swin Transformer. arXiv:2108.10257 [cs, eess].

Lugmayr, A., Danelljan, M., Timofte, R., 2019. Unsupervised Learning for Real-World Super-Resolution. arXiv:1909.09629 [cs, eess].

Rohith, G., Kumar, L.S., 2021. Paradigm shifts in super-resolution techniques for remote sensing applications. The Visual Computer 37, 1965–2008.

Shermeyer, J., Van Etten, A., 2019. The Effects of Super-Resolution on Object Detection Performance in Satellite Imagery. arXiv:1812.04098 [cs].

Wang, X., Xie, L., Dong, C., Shan, Y., 2021. Real-ESRGAN: Training Real-World Blind Super-Resolution with Pure Synthetic Data. arXiv:2107.10833 [cs, eess].

White, J., Codoreanu, A., Zuleta, I., Lynch, C., Marchisio, G., Petrie, S., Duffy, A.R., 2021. Super-Resolving Beyond Satellite Hardware Using Realistically Degraded Images. arXiv:2103.06270 [cs, eess].

Yalçin, İ., Kocaman Gökçeoğlu, S., Saunier, S., Albinet, C., 2021. Radiometric quality assessment for maxar hd imagery. 2021 24th ISPRS Congress 43. https://doi.org/10.5194/isprs-archives-xliii-b3-2021-797-2021

Yang, W., Zhang, X., Tian, Y., Wang, W., Xue, J.-H., Liao, Q., 2019. Deep learning for single image super-resolution: A brief review. IEEE Transactions on Multimedia 21, 3106–3121.

Zhang, K., Li, Y., Zuo, W., Zhang, L., Van Gool, L., Timofte, R., 2021a. Plug-and-Play Image Restoration with Deep Denoiser Prior. arXiv:2008.13751 [cs, eess].

Zhang, K., Liang, J., Van Gool, L., Timofte, R., 2021b. Designing a Practical Degradation Model for Deep Blind Image Super-Resolution. arXiv:2103.14006 [cs, eess].

Zhang, L., Dong, R., Yuan, S., Li, W., Zheng, J., Fu, H., 2021. Making Low-Resolution Satellite Images Reborn: A Deep Learning Approach for Super-Resolution Building Extraction. Remote Sensing 13, 2872. https://doi.org/10.3390/rs13152872

Zhang, R., Isola, P., Efros, A.A., Shechtman, E., Wang, O., 2018. The Unreasonable Effectiveness of Deep Features as a Perceptual Metric. arXiv:1801.03924 [cs].

Zhang, Y., Li, Kunpeng, Li, Kai, Wang, L., Zhong, B., Fu, Y., 2018. Image super-resolution using very deep residual channel attention networks, in: Proceedings of the European Conference on Computer Vision (ECCV). pp. 286–301.

Zhu, X., Talebi, H., Shi, X., Yang, F., Milanfar, P., 2020. Super-Resolving Commercial Satellite Imagery Using Realistic Training Data. arXiv:2002.11248 [cs, eess].