EUROPEAN CONFERENCE ON COMPUTER VISION
MILANO 2024

UNIVERSITY OF MANNHEIM

MAX PLANCK INSTITUTE FOR INFORMATICS

Fraunhofer ITWM

Improving Feature Stability during Upsampling – Spectral Artifacts and the Importance of Spatial Context



Shashank Agnihotri   Julia Grabinski   Prof. Dr.-Ing Margret Keuper

Motivation

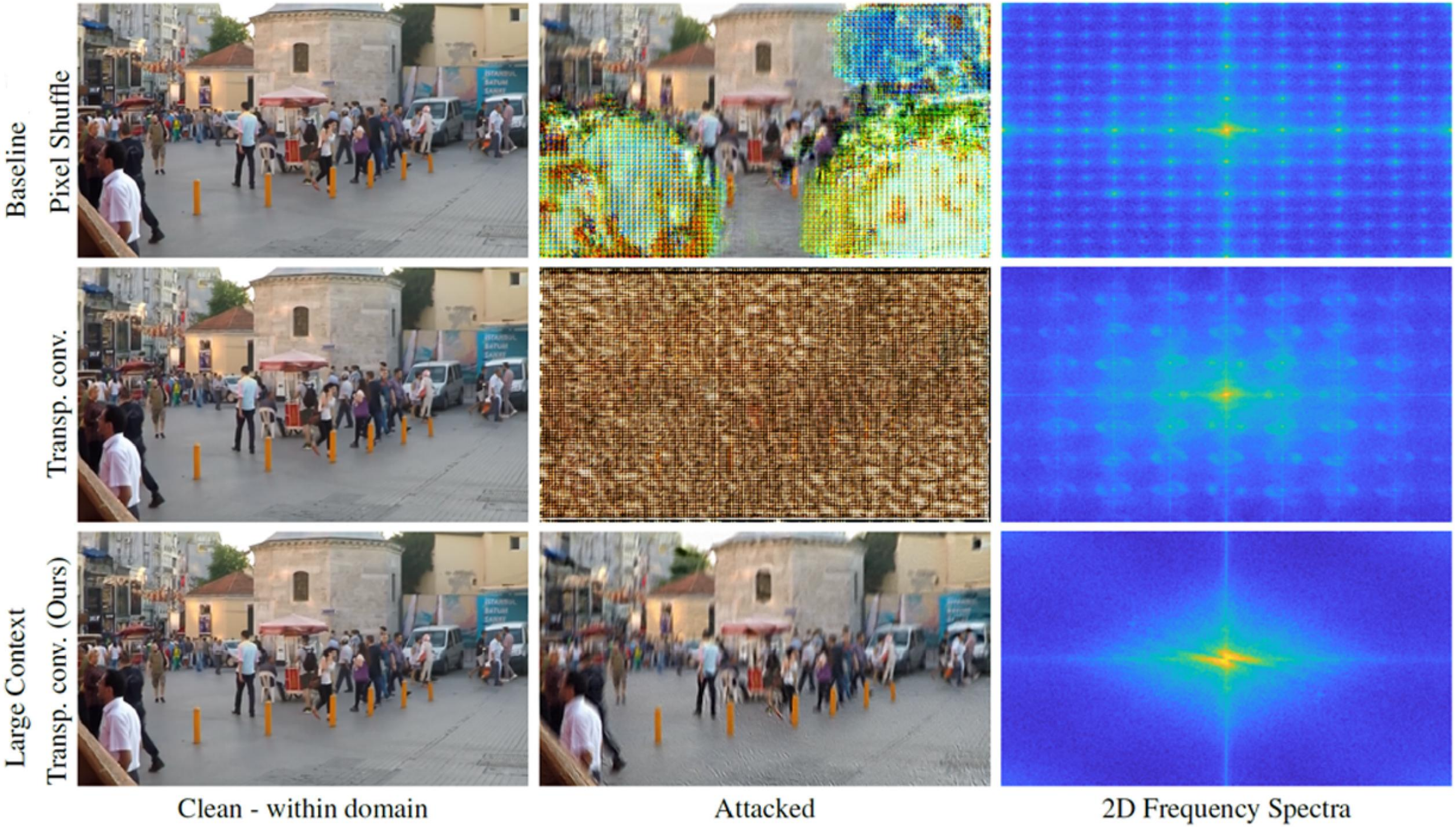

Figure 1. Image restoration using NAFNet[1] variants on GoPro images[2]

Introduction

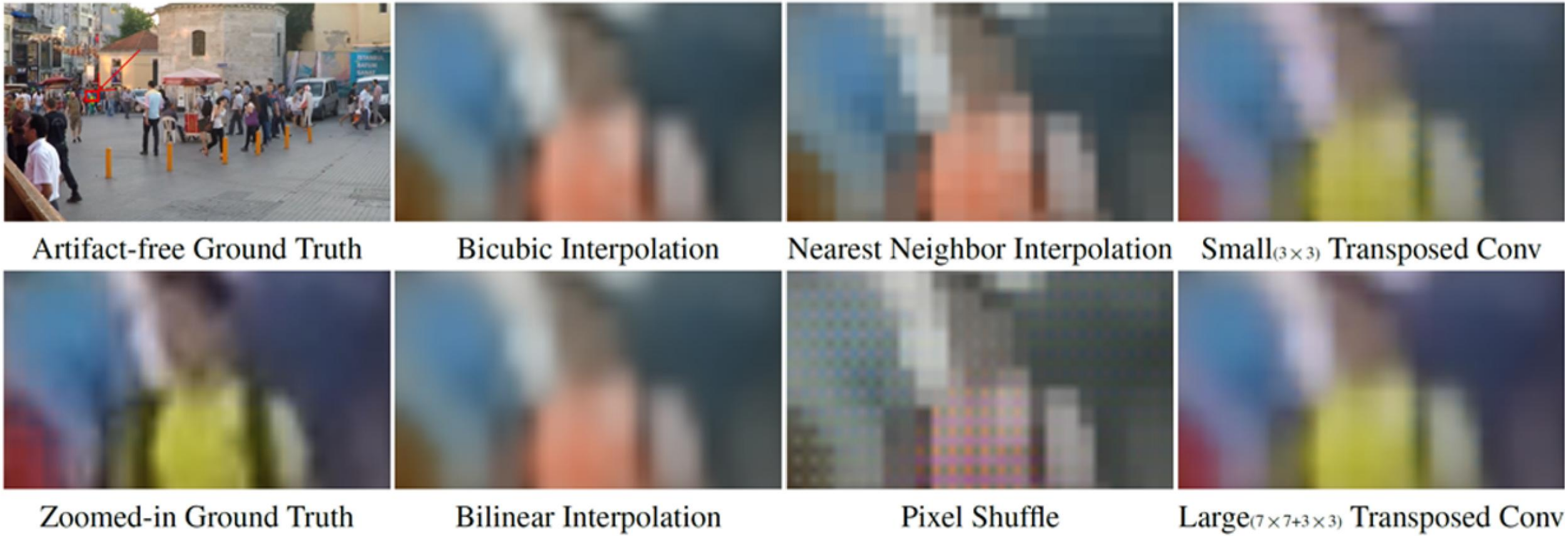

Figure 2. Images downsampled with 3x3 MaxPooling and then upsampled using various upsampling techniques.

What causes spectral artifacts?

Consider, w.l.o.g. a 1D signal *I* and its DFT $\mathcal{F}(I)$ with *k* being the index of discrete frequencies,

$$\boxed{\mathcal{F}(I)^{\text{up}}_{\bar{k}} = \sum_{j=0}^{2N-1} e^{-2\pi i \cdot \frac{j\bar{k}}{2 \cdot N}} \cdot I^{\text{up}}_{j} = \sum_{j=0}^{N-1} e^{-2\pi i \cdot \frac{2 \cdot j\bar{k}}{2 \cdot N}} I_j + \sum_{j=0}^{N-1} e^{-2\pi i \cdot \frac{2 \cdot (j+1)\bar{k}}{2 \cdot N}} \bar{I}_j, \qquad (1)}$$

What causes spectral artifacts?

Consider, w.l.o.g. a 1D signal I and its DFT $\mathcal{F}(I)$ with k being the index of discrete frequencies,

$$\mathcal{F}(I)^{\text{up}}_{\bar{k}} = \sum_{j=0}^{2N-1} e^{-2\pi i \cdot \frac{j\bar{k}}{2 \cdot N}} \cdot I^{\text{up}}_j = \sum_{j=0}^{N-1} e^{-2\pi i \cdot \frac{2 \cdot j\bar{k}}{2 \cdot N}} I_j + \sum_{j=0}^{N-1} e^{-2\pi i \cdot \frac{2 \cdot (j+1)\bar{k}}{2 \cdot N}} \bar{I}_j, \qquad (1)$$

During decoding, we upsample I to I^{up} with a factor of 2, for $\bar{k} = 0, \ldots, 2N-1$

$$\boxed{(1) = \sum_{j=0}^{2N-1} e^{-2\pi i \cdot \frac{j\bar{k}}{2 \cdot N}} \cdot \sum_{t=-\infty}^{\infty} I^{\text{up}}_j \cdot \delta(j - 2t). \qquad (2)}$$

What causes spectral artifacts?

Consider, w.l.o.g. a 1D signal *I* and its DFT $\mathcal{F}(I)$ with *k* being the index of discrete frequencies,

$$\mathcal{F}(I)^{\mathrm{up}}_{\bar{k}} = \sum_{j=0}^{2N-1} e^{-2\pi i \cdot \frac{j\bar{k}}{2\cdot N}} \cdot I^{\mathrm{up}}_j = \sum_{j=0}^{N-1} e^{-2\pi i \cdot \frac{2\cdot j\bar{k}}{2\cdot N}} I_j + \sum_{j=0}^{N-1} e^{-2\pi i \cdot \frac{2\cdot (j+1)\bar{k}}{2\cdot N}} \bar{I}_j, \tag{1}$$

During decoding, we upsample *I* to I^{up} with a factor of 2, for $\bar{k}$ = 0, ..., 2*N*-1

$$(1) = \sum_{j=0}^{2N-1} e^{-2\pi i \cdot \frac{j\bar{k}}{2\cdot N}} \cdot \sum_{t=-\infty}^{\infty} I^{\mathrm{up}}_j \cdot \delta(j - 2t). \tag{2}$$

the second term in (1) can be dropped in bed of nails sampling where $\bar{I}_j$=0
Since, the first term resembles $\mathcal{F}(I)$, we rewrite (1) using a Dirac impluse comb as,

$$\mathcal{F}(I)^{\mathrm{up}}_{\bar{k}} = \frac{1}{2} \sum_{t=-\infty}^{\infty} \left(\sum_{j=-\infty}^{\infty} e^{-2\pi i \cdot \frac{j\bar{k}}{2N}} I^{\mathrm{up}}_j \right) \left(\bar{k} - \frac{t}{2} \right) \tag{3}$$

$$\stackrel{(1)}{=} \frac{1}{2} \sum_{t=-\infty}^{\infty} \left(\sum_{j=-\infty}^{\infty} e^{-2\pi i \cdot \frac{j\bar{k}}{N}} \cdot I_j \right) \left(\bar{k} - \frac{t}{2} \right) = \frac{1}{2} \sum_{t=-\infty}^{\infty} \mathcal{F}(I)_{\bar{k}} \left(\bar{k} - \frac{t}{2} \right).$$

What causes spectral artifacts?

$$\mathcal{F}(I)^{\text{up}}_{\bar{k}} \overset{(1)}{=} \frac{1}{2}\sum_{t=-\infty}^{\infty}\left(\sum_{j=-\infty}^{\infty} e^{-2\pi i\cdot\frac{j\bar{k}}{N}}\cdot I_j\right)\left(\bar{k}-\frac{t}{2}\right) = \frac{1}{2}\sum_{t=-\infty}^{\infty}\mathcal{F}(I)_{\bar{k}}\left(\bar{k}-\frac{t}{2}\right). \qquad (3)$$

Such upsampling creates high-frequency replica of the signal at $t/2$ for t in $-\infty, \ldots, \infty$ in $\mathcal{F}(I)^{\text{up}}$.

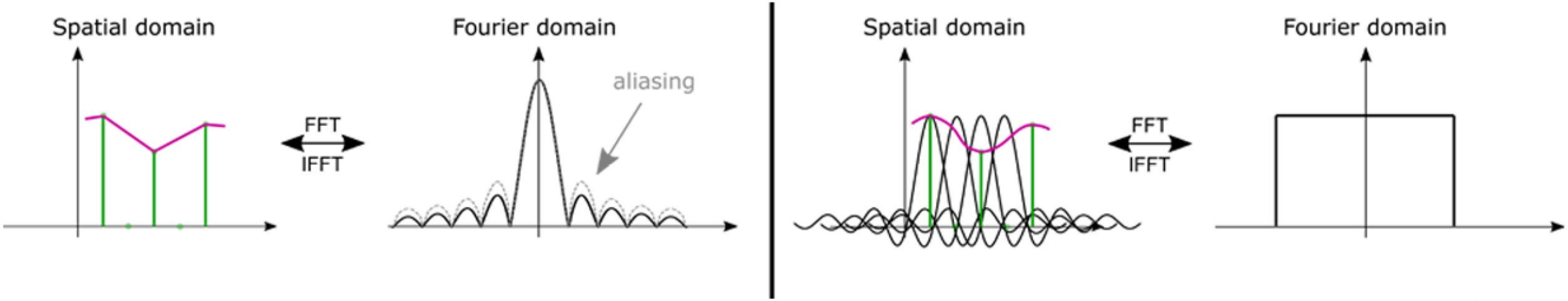

Figure 3. (Left) Linear interpolation (pink) of the sample (green) causes aliases. (Right) Optimal signal reconstruction (pink) is achieved by sinc interpolation.

Hypothesis

***Hypothesis 1**: Large Context Transposed Convolutions (LCTC)* **i.e. Large kernels in transposed convolution operations provide more context and reduce spectral artifacts and can therefore be leveraged by the network to facilitate better and more robust pixel-wise predictions.**

Hypothesis

Hypothesis 1: Large Context Transposed Convolutions (LCTC) **i.e. Large kernels in transposed convolution operations provide more context and reduce spectral artifacts and can therefore be leveraged by the network to facilitate better and more robust pixel-wise predictions.**

Hypothesis 2 (Null Hypothesis H2): **To leverage prediction context and reduce spectral artifacts, it is crucial to increase the size of the transposed convolution kernels (upsample using large filters). Increasing the size of normal (i.e. non-upsampling) decoder convolutions does not have this effect.**

Method

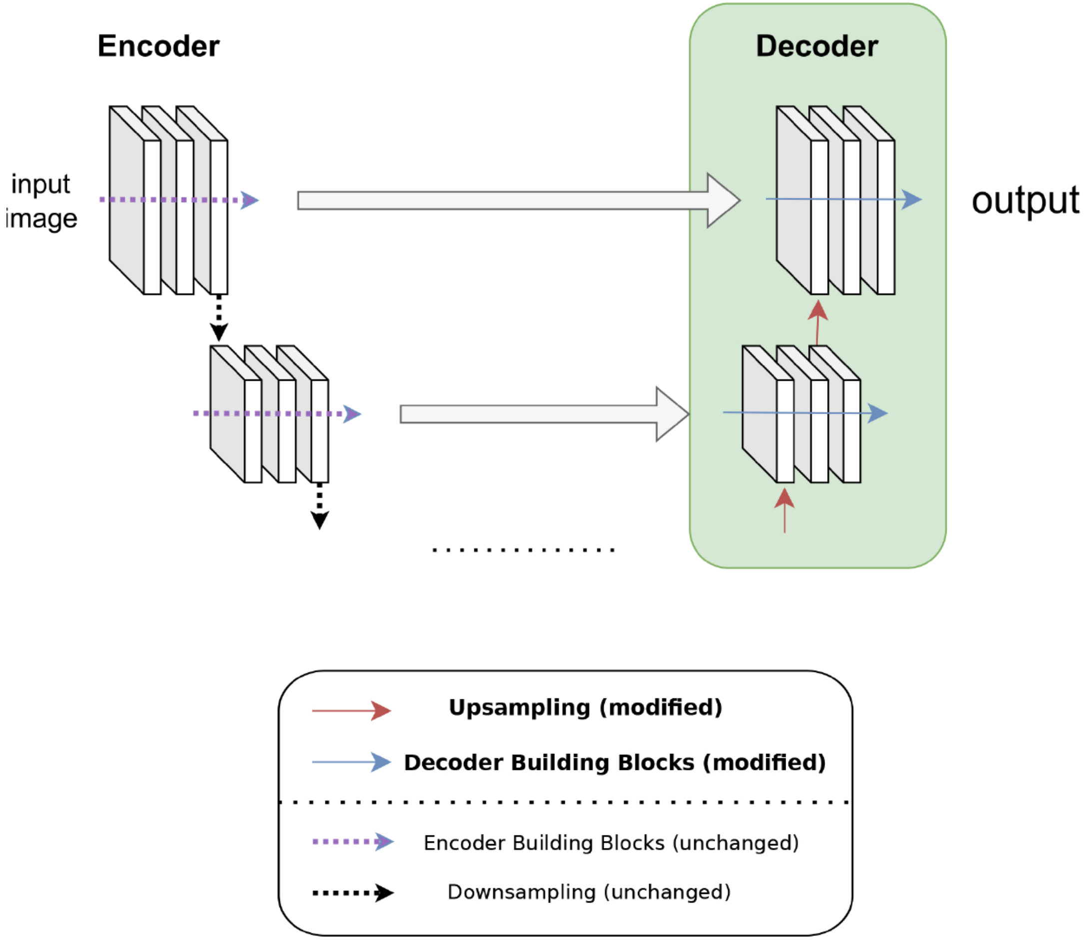

Figure 4. Abstract representation of the encoder-decoder architectural modifications. Our study focuses on the model decoder (in green).

Method

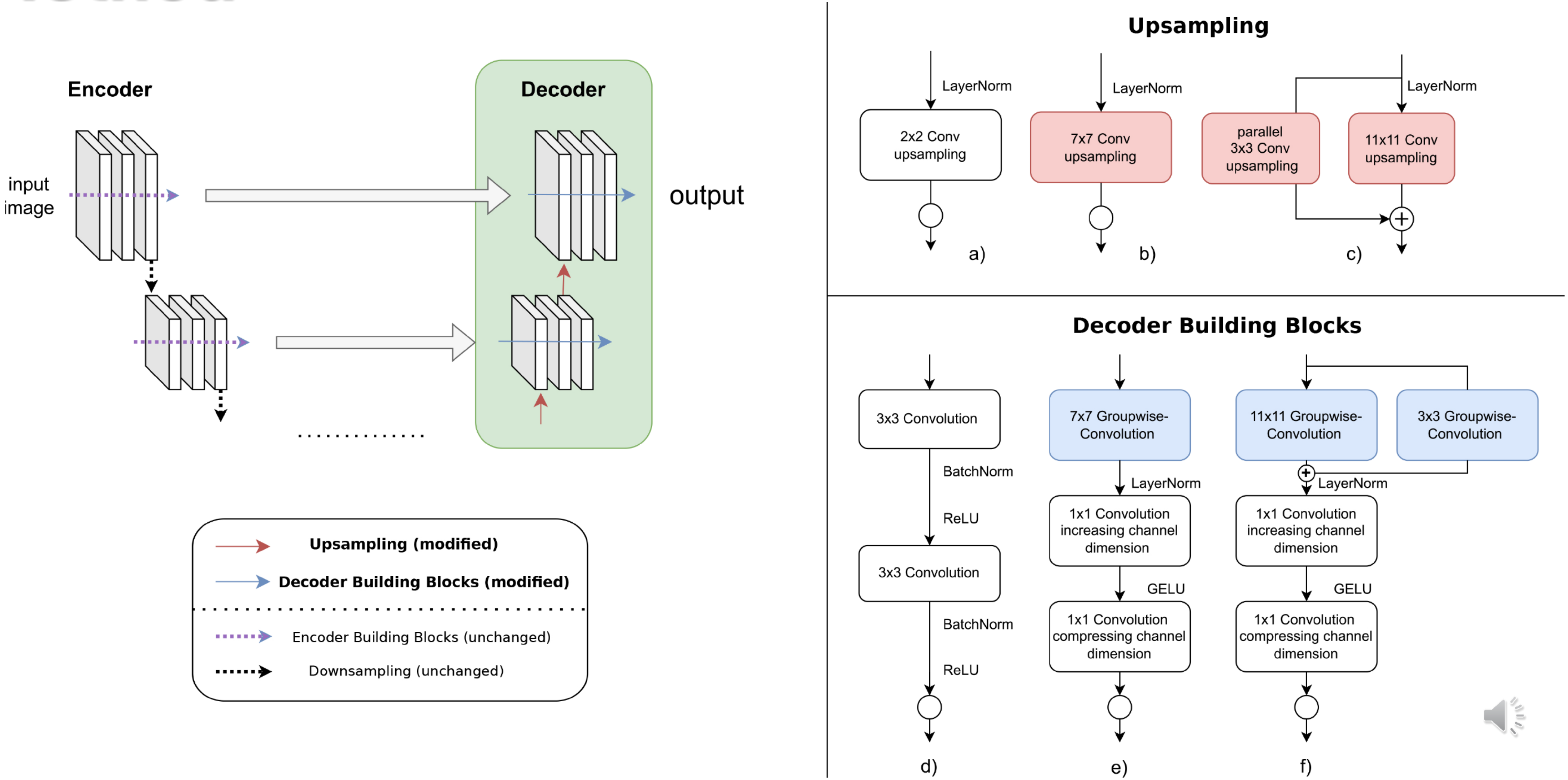

Figure 4. Abstract representation of the encoder-decoder architectural modifications. Our study focuses on the model decoder (in green).

Results

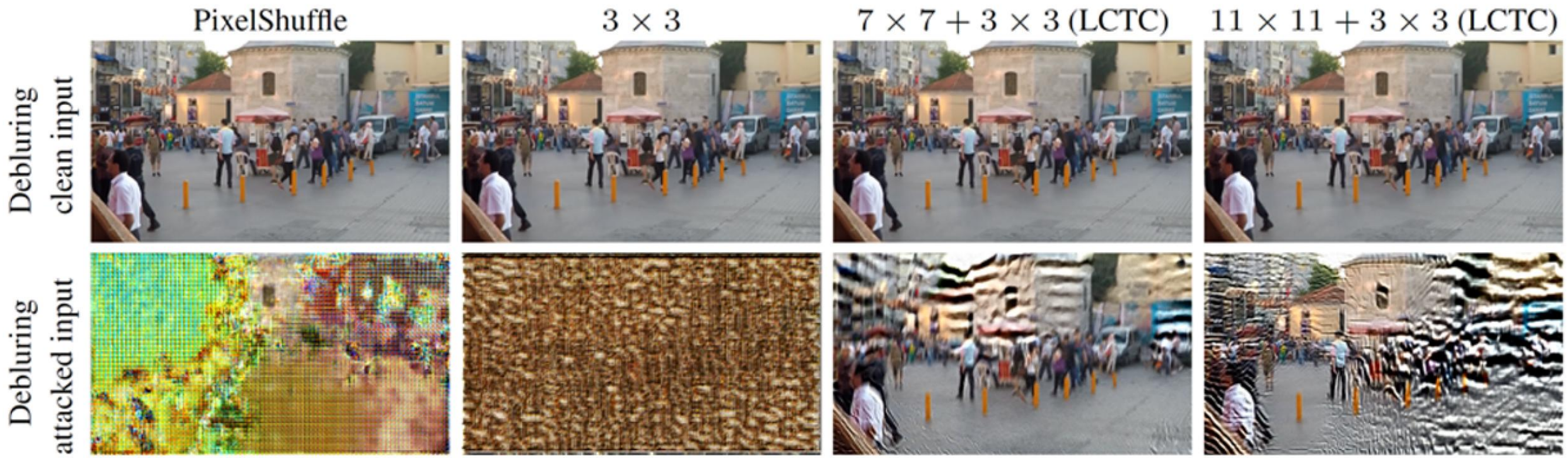

Figure 5. Restorations using NAFNet (uses PixelShuffle) and its variants including LCTC under PGD attack.

Results

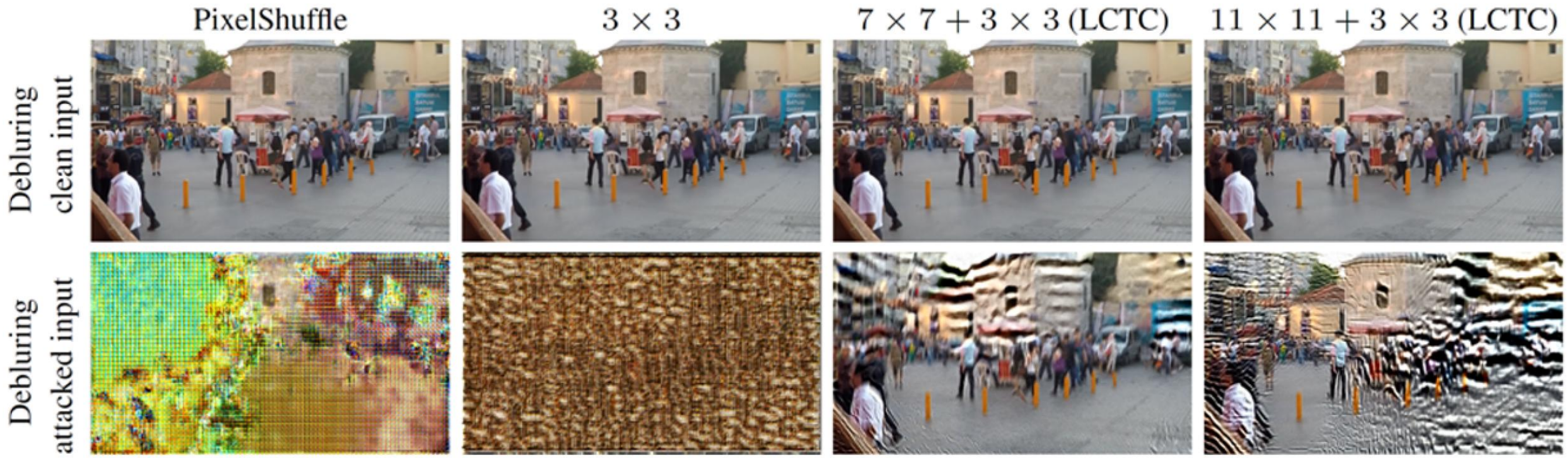

Figure 5. Restorations using NAFNet (uses PixelShuffle) and its variants including LCTC under PGD attack.

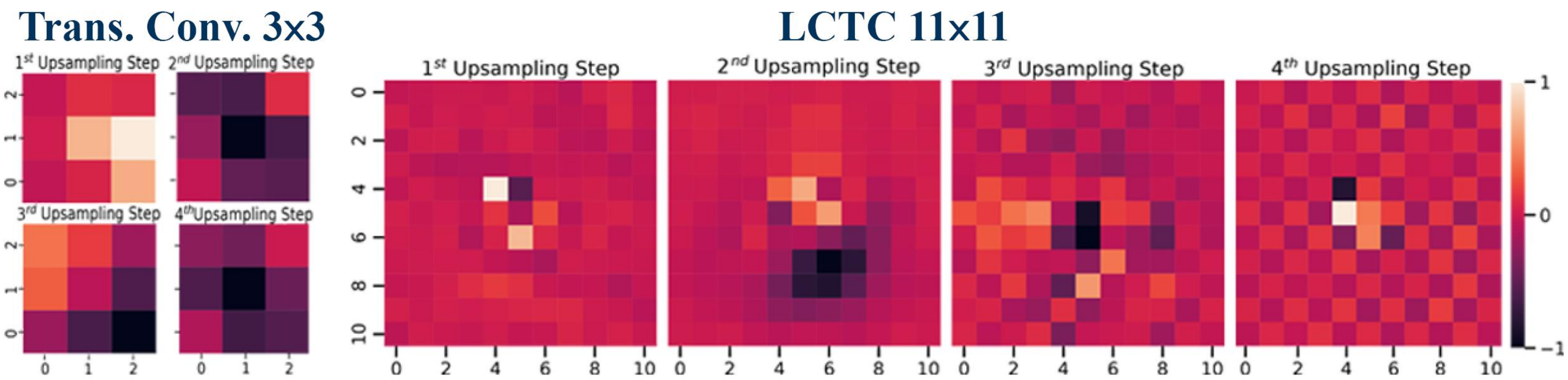

Figure 6. Normalized kernel weights from a random channel for the NAFNet models above.

Results

Figure 7. Difference in semantic segmenation mask predictions between LCTC and baseline transposed convolution using UNet[3] with ConvNeXt[4] encoder on PASCAL VOC2012[5] image

Results

Figure 7. Difference in semantic segmenation mask predictions between LCTC and baseline transposed convolution using UNet[3] with ConvNeXt[4] encoder on PASCAL VOC2012[5] image

Table 1. Quantitative results for the above models.

Transposed Convolution Kernels	**Clean Test Accuracy**			**SegPGD ($\epsilon \approx \frac{8}{255}$) attack iterations** 3			20		
	mIoU	mAcc	allAcc	mIoU	mAcc	allAcc	mIoU	mAcc	allAcc
2×2 (baseline)	78.34	86.89	95.15	23.06	46.51	45.30	5.54	18.79	23.72
LCTC: 7×7 (Ours)	78.92	**88.06**	95.23	26.53	53.05	61.16	**7.17**	23.05	**27.52**
LCTC: 11×11 + 3×3 (Ours)	**79.33**	87.81	**95.41**	**27.49**	**53.08**	**64.13**	7.08	**23.30**	26.82

Conclusion

- **We provide conclusive reasoning and empirical evidence for our hypotheses on the importance of context during data upsampling.**
- **Increasing the size of convolutions upsampling (LCTC) increases prediction stability.**
- **Increasing the size of those convolution layers without upsampling does not benefit the network.**
- **We show that observations made for increased context during encoding do not translate to decoding.**
- **Large Context Transposed Convolutions can be directly incorporated into recent models.**

References

[1] Chen, L., Chu, X., Zhang, X., Sun, J. (2022). Simple Baselines for Image Restoration. In: Avidan, S., Brostow, G., Cissé, M., Farinella, G.M., Hassner, T. (eds) Computer Vision – ECCV 2022. ECCV 2022. Lecture Notes in Computer Science, vol 13667. Springer, Cham. https://doi.org/10.1007/978-3-031-20071-7_2

[2] Nah, S., Hyun Kim, T., & Mu Lee, K. (2017). Deep multi-scale convolutional neural network for dynamic scene deblurring. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 3883-3891).

[3] Ronneberger, Olaf, Philipp Fischer, and Thomas Brox. "U-Net: Convolutional networks for biomedical image segmentation." Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18. Springer International Publishing, 2015.

[4] Liu, Z., Mao, H., Wu, C. Y., Feichtenhofer, C., Darrell, T., & Xie, S. (2022). A convnet for the 2020s. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 11976-11986).

[5] M. Everingham, L. Van Gool, C. Williams, J. Winn, and A. Zisserman. http://www.pascal-network.org/challenges/VOC/voc2012/workshop/index.html, (2012)

EUROPEAN CONFERENCE ON COMPUTER VISION

MILANO 2024

UNIVERSITY OF MANNHEIM

MAX PLANCK INSTITUTE FOR INFORMATICS

Fraunhofer ITWM

Improving Feature Stability during Upsampling – Spectral Artifacts and the Importance of Spatial Context

Shashank Agnihotri

Julia Grabinski

Prof. Dr.-Ing Margret Keuper

Paper: