Improvements on Deep Fourier Up-Sampling

Hefeng Zhou
University of Michigan
EECS 556
zhouhf@umich.edu

Yiwei Gui
University of Michigan
EECS 556
yiweigui@umich.edu

Yiting Wang
University of Michigan
EECS 556
yitinw@umich.edu

Jiadong Hao
University of Michigan
EECS 556
haojd@umich.edu

Honor Code: Our team will give attribution for any figures used in our documents and will cite all code sources (beyond those already built into Julia Base or Matlab toolboxes).

1 Introduction

1.1 Overviews

Nowadays, spatial up-sampling is extensively adopted in multi-scale modeling approaches (e.g., U-Net, feature pyramids). Traditional spatial up-sampling operators such as interpolation and transposed convolutions heavily rely on local pixel correlations, which often lead to blurred edges or lost high-frequency details. In contrast, Fourier domain inherently supports global modeling, which provides a brand new area for up-sampling. However, up-sampling in the frequency domain faces challenges due to the absence of invariant properties and local texture similarity; using naive interpolation in the spatial domain will cause errors.

1.2 Solutions and Advantages

Solutions The paper "Deep fourier Up-sampling"(1) developed FourierUp, which is a learnable operator that can be directly integrated into existing neural networks. It works by first transforming a low-resolution feature map from the spatial domain into the Fourier domain using a 2D discrete Fourier transform. Once in the Fourier domain, specialized upsampling techniques—such as 1)periodic padding, 2)area interpolation, or 3)corner interpolation—are applied to increase the resolution of the frequency representation according to theoretical transform rules verified in the paper(1). Finally, an inverse Fourier transform is performed to convert the enhanced frequency map back into the spatial domain, yielding a high-resolution feature map that retains both the global context and the fine details of the original image. This approach not only overcomes the limitations of conventional upsampling methods but also provides a robust, multi-scale solution that can be integrated into various deep neural network architectures.

Advantages FourierUp offers a powerful alternative to conventional spatial up-sampling by leveraging the global modeling capabilities inherent to the Fourier domain. Unlike traditional methods that rely on local pixel relationships Fourier up-sampling operates on a global scale, it captures long-range dependencies and multi-scale frequency patterns through the spectral convolution theorem. This enables a more robust integration of features across different resolutions, which is particularly beneficial for complex computer vision tasks that demand both precise local detail and coherent global context. Furthermore, while previous approaches have limited the interaction between spatial and Fourier representations to a single resolution scale, Deep Fourier Up-Sampling extends this capability across multiple scales. This multi-scale fusion enhances the recovery of fine textures and object boundaries and thus improves overall performance. Also, FourierUp can be easily integrated into existing networks and improves performance across diverse applications.

2 Quantitative Performance Prediction

Stage 1 We try to reproduce the results of incorporating the FourierUp into LPNet (2), which is a representative architecture for image de-raining. Aligned with the original paper (1), we use both Rain200H and Rain200L (3) as the training and testing sets, use Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM) as the evaluation metrics, and we expect to have the results shown in Fig.1, where the four model settings are defined below:

(1) Original: the baseline without any changes;

(2) FourierUp-AreaUp: replacing the original model's spatial up-sampling with the union of the Area-Interpolation variant of our FourierUp and the spatial up-sampling itself;

(3) FourierUp-Padding: replacing the original model's spatial up-sampling operator with the union of the Periodic-Padding variant of our FourierUp and the spatial up-sampling itself;

(4) Spatial-Up: replacing the variants of FourierUp in configurations 2 and 3 with spatial up-sampling.

Model	Configurations	Rain200H		Rain200L	
		PSNR	SSIM	PSNR	SSIM
LPNet	Original	22.907	0.775	32.461	0.947
	Spatial-Up	22.956	0.777	32.522	0.950
	FourierUp-AreaUp	22.163	0.783	32.681	0.954
	FourierUp-Padding	**23.295**	**0.786**	**32.835**	**0.956**

Figure 1: Performance comparison of LPNet with different up-sampling methods.

Then, we want to improve the FourierUp module and re-evaluate it on the same task. Hence, the modified version is expected to have higher PSNR and SSIM than the "FourierUp-Padding LPNet" mentioned in the paper(1).

Stage 2 We extend our modified FourierUp module to the image segmentation task to verify its universality. We choose the baseline model to be the DeepLabv3 (4) and incorporate FourierUp module into it. To evaluate our modified DeepLab-FourierUp, we choose Cityscapes(5) dataset which consists of 2975, 500, and 1525 samples for the training, validation, and test sets respectively. We select Mean Intersection over Union (mIoU) as the evaluation metric, which measures the overlap between the predicted segmentation and ground truth for each class, averaged over all classes. Based on our integration of deep Fourier Up-Sampling via a fusion module, we anticipate at least 3% improvement over the baseline DeepLabV3+ that achieves 81.3%.

3 Plan

3.1 Stage 1: Reimplementation and Improvement of FourierUp

3.1.1 Preparation

Download dataset Rain200H and Rain200L from `https://github.com/csdwren/PReNet`

Download the code of LPNet from `https://github.com/manman1995/Deep-Fourier-Upsampling/tree/main/deraining/lpnet`

Install the required environment.

3.1.2 Reproduction and Comparison

Reproduce the results of LPNet with FourierUp.

Compare the results gained by directly applying the hard math formula of FourierUp and the learnable FourierUp.

3.1.3 Modification

Modify the code to test the impact of larger size of kernel to output.

Modify the code to combine the periodic-padding variant and the area interpolation variant.

3.1.4 Evaluation

Evaluate FourierUp with larger kernel sizes, and hybrid FourierUp implementation on Rain200H and Rain200L with PSNR and SSIM and compare to the original model.

3.2 Stage 2: Application of the Improved FourierUp to Image Segmentation

3.2.1 Preparation

Download dataset Cityscapes from `https://www.cityscapes-dataset.com/`

Download the code of DeepLabv3 from `https://github.com/leimao/DeepLab-V3`

Install the required environment.

3.2.2 Modification

Modify the code to incorporate the modified FourierUp module into DeepLabv3.

3.2.3 Evaluation

Evaluate the new architecture DeepLab-FourierUp on CityScapes dataset with Mean Intersection over Union (mIoU) and compare to the original model.

3.3 Work allocation

- **Yiwei Gui:** Reproduce the results of LPNet with FourierUp on Rain200H and Rain200L datasets. Reproduce the results of DeepLabv3 on image segmentation dataset "Cityscapes".
- **Hefeng Zhou:** Make a comparision evaluation on "pure-math-formula-based" DeepFourier upsampling vs "learnable" DeepFourier upsampling.
- **Yiting Wang:** Propose new ways of learning the FourierUp module and use LPNet to verify the potential improvements.
- **Jiadong Hao:** Integrate the improved FourierUp module with DeepLabv3 on "Cityscapes" to verify that our proposed up-sampling module can be generally used in a wide range of image processing tasks and gain improvements.

4 Extensions

4.1 Modification of FourierUp with Larger Kernel Size

The Fourier transform operates globally, meaning that each point in the Fourier spectrum is influenced by information from the entire input image. By incorporating larger convolution kernels, the system can become more robust to noise. Additionally, they function as low-pass filters, mitigating interpolation artifacts—particularly in high-frequency regions—resulting in cleaner and more stable outputs.

In the original FourierUp method, a learnable 1×1 convolution is applied after periodic padding in the Fourier domain, rather than directly using the 1/4 scaling factor derived from Theorem-1. To further analyze the impact of kernel size, we will experiment with larger convolution kernels, starting with 3×3 and 5×5. The evaluation will focus on accuracy, run-time efficiency, and model stability to understand how kernel size influences the overall performance of the FourierUp method.

4.2 A hybrid approach to FourierUp implementation

The original project introduced two distinct up-sampling variants: the periodic-padding variant and the area interpolation variant. The periodic-padding variant, while effective in maintaining structural consistency, is prone to introducing boundary artifacts. On the other hand, the area-interpolation variant, though better at preserving high-frequency details, may result in the distortion of fine image structures.

Our proposed method directly combines the Periodic-Padding variant and Area-Interpolation variant during the implementation of FourierUp. The results from this combined approach, along with the results from conventional spatial up-sampling, serve as inputs to a final integration stage. This three-way combination of results (Periodic-Padding FourierUp, Area-Interpolation FourierUp, and spatial up-sampling) is then synthesized to produce the final output. This approach allows us to leverage the complementary strengths of each method in a unified framework.

4.3 DeepLab-FourierUp

To verify the versality of the improved Deep Fourier up-sampling module, we aim to apply it to image segmentation, an image processing task that is not discussed in the original paper. The baseline model we choose is DeepLabV3 (1), whose architecture is shown in Fig.2.

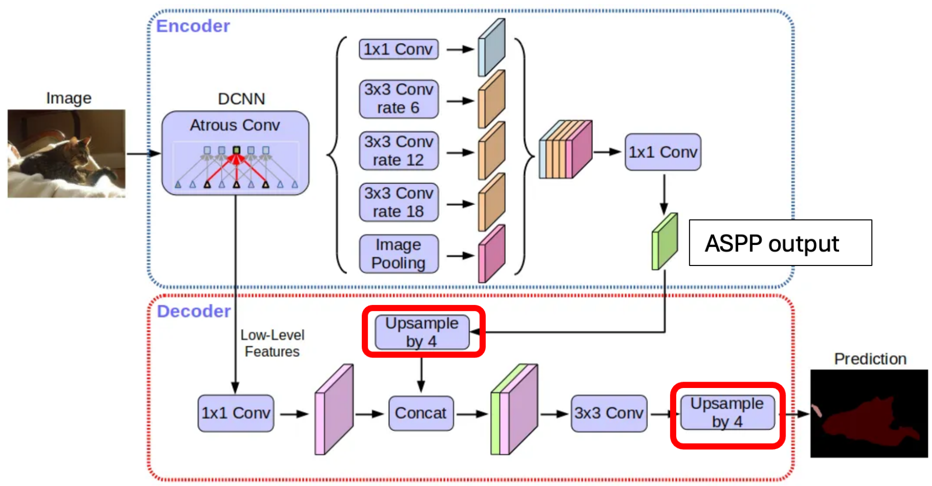

Figure 2: DeepLabV3 Architecture

In the standard design, the DeepLabv3 decoder uses bilinear interpolation to upscale feature maps, typically in the following two main stages:

1. Upsampling ASPP Outputs: To match the spatial resolution of low-level features for fusion.
2. Final Upsampling: To bring the fused feature map to the original input resolution.

To take advantage of spatial and frequency domain upsampling, we will adpot a fusion strategy that combines:

1. Conventional Spatial Upsampling in DeepLabv3: (e.g., bilinear interpolation), which is efficient and preserves local details.
2. Deep Fourier Up-Sampling: which operates in the frequency domain to capture global context and improve high-frequency detail reconstruction.

The final up-sampled output is obtained by computing a weighted sum of the two branches shown in Eq.(1). A learnable parameter (or set of parameters) controls the balance between the spatial and Fourier outputs, enabling the network to adapt during training.

$$\text{Output} = \alpha \times \text{SpatialUpsampled} + (1 - \alpha) \times \text{FourierUpsampled} \tag{1}$$

References

[1] M. Zhou, H. Yu, J. Huang, F. Zhao, J. Gu, C. C. Loy, D. Meng, and C. Li, "Deep fourier up-sampling," *Advances in Neural Information Processing Systems*, vol. 35, pp. 22995–23008, 2022.

[2] X. Fu, B. Liang, Y. Huang, X. Ding, and J. Paisley, "Lightweight pyramid networks for image deraining," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 31, no. 6, pp. 1794–1807, 2020.

[3] D. Ren, W. Zuo, Q. Hu, P. Zhu, and D. Meng, "Progressive image deraining networks: A better and simpler baseline," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3937–3946, June 2019.

[4] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking atrous convolution for semantic image segmentation," *arXiv preprint arXiv:1706.05587*, 2017.

[5] M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, and B. Schiele, "The cityscapes dataset for semantic urban scene understanding," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3213–3223, 2016.