

---

# QwinSR: An All-MLP Shifted Window Model for Image Super Resolution

---

**Mark Bauer\***  
Siebel Department of Computer Science  
University of Illinois Urbana-Champaign  
markb6@illinois.edu

**Quinn Ouyang†**  
School of Music  
University of Illinois Urbana-Champaign  
qouyang3@illinois.edu

## Abstract

Since its inception, the Swin Transformer backbone architecture has consistently showcased remarkable performance across a range of well-established computer vision benchmarks. To enhance computational efficiency, the Swin Mixer architecture has adopted a similar structure. However, it distinguishes itself by substituting transformers with Mixer Layers, thus giving rise to Swin Mixer Layers (SMLs). In line with this design, we introduce QwinSR, an application of this all-MLP architecture tailored for the task of single image super-resolution. This adaptation simplifies the original transformer-based image restoration model, SwinIR. QwinSR leverages SMLs to extract essential features and subsequently aggregates them within a compact convolutional neural network, facilitating image reconstruction. We anticipate that this approach will yield competitive accuracy-to-computation ratios, particularly when compared to SwinIR and other leading models in the field.

## 1 Introduction

Despite advances in modern photography and image transmission technology, resolution loss is often an unavoidable or necessary compromise. This produces a need and desire for techniques to construct higher fidelity images from lower resolution sources, which we call super resolution (SR) imaging. However, SR is a more niche category than other computer vision tasks (such as classification and semantic segmentation []), making this field ripe for new research. Given the simple objective and need for research on SR, we propose a simplified all-MLP model based on the shifted window design from Swin Transformer: QwinIR.

### 1.1 Related Work

#### 1.1.1 Single Image Super-Resolution

“Super-resolution” generally refers to the process of enhancing the visual detail and fidelity of an image by predicting pixels to increase its resolution []. In other literature, SR is roughly interchangeable with the more general terms “upsampling” and “reconstruction.” Note that the latter implies that an exact / true higher resolution image exists for a lower resolution one, which typically comes from artificial downsampling (classic image SR) []. “Single image” differentiates from the other common variant of this task, multiple image SR, which assumes the advantage of having more than one source image. Effective techniques exploit the related images for more pixel information to construct from, which can apply in the real-world (e.g. a burst shot, video frames, etc.) []. However, we focus on single image SR as it is a more fundamental and challenging task.

---

\*Use footnote for providing further information about author (webpage, alternative address)—*not* for acknowledging funding agencies.

†“Quinn” = “Qwin”



Figure 1: Visual comparisons of traditional and learning-based classic image SR approaches to a  $\times 4$ -upscaled original high resolution image.

Traditional algorithmic approaches for SR directly interpolate pixels from a lower resolution source, typically assuming a downsampling process to generalize the reconstruction. Popular basic algorithms include bicubic and nearest-neighbor interpolation which are fast and eschew the long training times associated with learning-based models, but the absence of trained priors obviously limits their ability to hallucinate new pixels [1].

### 1.1.2 Learning-based Approaches

On the other hand, learning-based models spanning a variety of architectures have far surpassed the generalized interpolations that traditional approaches limit themselves to [1], as we have briefly illustrated in Figure 1. Models based on convolutional neural networks (e.g. SRCNN, SRResNet), general adversarial networks (e.g. ESRGAN, SRGAN, Real-ESRGAN), and transformers (e.g. SwinIR) have all effectively competed for state-of-the-art performance [2].

For SR related tasks, CNNs are effective at global feature extraction but have lately been usurped by transformers, which tend to be more robust and lightweight. As a consequence, pure CNN models often require added complexity just to match the performance of transformer models [https://arxiv.org/pdf/2105.01601.pdf]. It is quite common to utilize hybrid CNN-transformer models, in which CNNs are used for global feature extraction while transformers focus on local feature extraction [3]. E.g. ...

GANs tend to be tedious to fine tune and also have many trainable parameters, making the training process very long [4]. Similarly, diffusion models like StableSR have gigantic datasets and are quite complex [5]. The focus of this research project is on simplicity, and as such we do not consider these models.

## 2 Citations, figures, tables, references

These instructions apply to everyone.

### 2.1 Citations within the text

The natbib package will be loaded for you by default. Citations may be author/year or numeric, as long as you maintain internal consistency. As to the format of the references themselves, any style is acceptable as long as it is used consistently.

The documentation for natbib may be found at

<http://mirrors.ctan.org/macros/latex/contrib/natbib/natnotes.pdf>

Of note is the command `\citet`, which produces citations appropriate for use in inline text. For example,

```
\citet{hasselmo} investigated\dots
```



Figure 2: Sample figure caption.

produces

Hasselmo, et al. (1995) investigated...

If you wish to load the `natbib` package with options, you may add the following before loading the `neurips_2023` package:

```
\PassOptionsToPackage{options}{natbib}
```

If `natbib` clashes with another package you load, you can add the optional argument `nonatbib` when loading the style file:

```
\usepackage[nonatbib]{neurips_2023}
```

As submission is double blind, refer to your own published work in the third person. That is, use “In the previous work of Jones et al. [4],” not “In our previous work [4].” If you cite your other papers that are not widely available (e.g., a journal paper under review), use anonymous author names in the citation, e.g., an author of the form “A. Anonymous” and include a copy of the anonymized paper in the supplementary material.

## 2.2 Footnotes

Footnotes should be used sparingly. If you do require a footnote, indicate footnotes with a number<sup>3</sup> in the text. Place the footnotes at the bottom of the page on which they appear. Precede the footnote with a horizontal rule of 2 inches (12 picas).

Note that footnotes are properly typeset *after* punctuation marks.<sup>4</sup>

## 2.3 Figures

All artwork must be neat, clean, and legible. Lines should be dark enough for purposes of reproduction. The figure number and caption always appear after the figure. Place one line space before the figure caption and one line space after the figure. The figure caption should be lower case (except for first word and proper nouns); figures are numbered consecutively.

You may use color figures. However, it is best for the figure captions and the paper body to be legible if the paper is printed in either black/white or in color.

## 2.4 Tables

All tables must be centered, neat, clean and legible. The table number and title always appear before the table. See Table 1.

---

<sup>3</sup>Sample of the first footnote.

<sup>4</sup>As in this example.

Table 1: Sample table title

Part		
Name	Description	Size ( $\mu\text{m}$ )
Dendrite	Input terminal	$\sim 100$
Axon	Output terminal	$\sim 10$
Soma	Cell body	up to $10^6$

Place one line space before the table title, one line space after the table title, and one line space after the table. The table title must be lower case (except for first word and proper nouns); tables are numbered consecutively.

Note that publication-quality tables *do not contain vertical rules*. We strongly suggest the use of the `booktabs` package, which allows for typesetting high-quality, professional tables:

<https://www.ctan.org/pkg/booktabs>

This package was used to typeset Table 1.

## 2.5 Math

Note that display math in bare TeX commands will not create correct line numbers for submission. Please use LaTeX (or AMSTeX) commands for unnumbered display math. (You really shouldn't be using  $\$$  anyway; see <https://tex.stackexchange.com/questions/503/why-is-preferable-to> and <https://tex.stackexchange.com/questions/40492/what-are-the-differences-between-align-equation-and-displaymath> for more information.)

## 2.6 Final instructions

Do not change any aspects of the formatting parameters in the style files. In particular, do not modify the width or length of the rectangle the text should fit into, and do not change font sizes (except perhaps in the **References** section; see below). Please note that pages should be numbered.

## 3 Preparing PDF files

Please prepare submission files with paper size "US Letter," and not, for example, "A4."

Fonts were the main cause of problems in the past years. Your PDF file must only contain Type 1 or Embedded TrueType fonts. Here are a few instructions to achieve this.

- You should directly generate PDF files using `pdflatex`.
- You can check which fonts a PDF file uses. In Acrobat Reader, select the menu Files>Document Properties>Fonts and select Show All Fonts. You can also use the program `pdf fonts` which comes with `xpdf` and is available out-of-the-box on most Linux machines.
- `xfig` "patterned" shapes are implemented with bitmap fonts. Use "solid" shapes instead.
- The `\bbold` package almost always uses bitmap fonts. You should use the equivalent AMS Fonts:

```
\usepackage{amsfonts}
```

followed by, e.g., `\mathbb{R}`, `\mathbb{N}`, or `\mathbb{C}` for  $\mathbb{R}$ ,  $\mathbb{N}$  or  $\mathbb{C}$ . You can also use the following workaround for reals, natural and complex:

```
\newcommand{\RR}{\mathbb{R}} %real numbers
\newcommand{\Nat}{\mathbb{N}} %natural numbers
\newcommand{\CC}{\mathbb{C}} %complex numbers
```

Note that `amsfonts` is automatically loaded by the `amssymb` package.

If your file contains type 3 fonts or non embedded TrueType fonts, we will ask you to fix it.

### 3.1 Margins in L<sup>A</sup>T<sub>E</sub>X

Most of the margin problems come from figures positioned by hand using `\special` or other commands. We suggest using the command `\includegraphics` from the `graphicx` package. Always specify the figure width as a multiple of the line width as in the example below:

```
\usepackage[pdftex]{graphicx} ...  
\includegraphics[width=0.8\linewidth]{myfile.pdf}
```

See Section 4.4 in the graphics bundle documentation (<http://mirrors.ctan.org/macros/latex/required/graphics/grfguide.pdf>)

A number of width problems arise when L<sup>A</sup>T<sub>E</sub>X cannot properly hyphenate a line. Please give LaTeX hyphenation hints using the `\-` command when necessary.

### Acknowledgments and Disclosure of Funding

Use unnumbered first level headings for the acknowledgments. All acknowledgments go at the end of the paper before the list of references. Moreover, you are required to declare funding (financial activities supporting the submitted work) and competing interests (related financial activities outside the submitted work). More information about this disclosure can be found at: <https://neurips.cc/Conferences/2023/PaperInformation/FundingDisclosure>.

Do **not** include this section in the anonymized submission, only in the final paper. You can use the `ack` environment provided in the style file to automatically hide this section in the anonymized submission.

## 4 Supplementary Material

Authors may wish to optionally include extra information (complete proofs, additional experiments and plots) in the appendix. All such materials should be part of the supplemental material (submitted separately) and should NOT be included in the main submission.

### References

References follow the acknowledgments in the camera-ready paper. Use unnumbered first-level heading for the references. Any choice of citation style is acceptable as long as you are consistent. It is permissible to reduce the font size to `small` (9 point) when listing the references. Note that the Reference section does not count towards the page limit.

- [1] Alexander, J.A. & Mozer, M.C. (1995) Template-based algorithms for connectionist rule extraction. In G. Tesauro, D.S. Touretzky and T.K. Leen (eds.), *Advances in Neural Information Processing Systems 7*, pp. 609–616. Cambridge, MA: MIT Press.
- [2] Bower, J.M. & Beeman, D. (1995) *The Book of GENESIS: Exploring Realistic Neural Models with the GEneral NEural Simulation System*. New York: TELOS/Springer-Verlag.
- [3] Hasselmo, M.E., Schnell, E. & Barkai, E. (1995) Dynamics of learning and recall at excitatory recurrent synapses and cholinergic modulation in rat hippocampal region CA3. *Journal of Neuroscience* **15**(7):5249-5262.