Waifu2x denoise guide for digital raws

*references to “noise” in this document refer to lossy compression artifacts

Downloads

For nvidia gpu with strictly greater than 2gb vram

- https://github.com/lltcggie/waifu2x-caffe/releases

For other gpu

- main program/CLI: https://github.com/nihui/waifu2x-ncnn-vulkan
- GUI for vulkan version: https://github.com/f11894/waifu2x-ncnn-vulkan-GUI/releases

Remember to download the releases and not the source code
For ncnn vulkan, use fp32 mode in gui preferences
side notes: AaronFeng753/Waifu2x-Extension-GUI does not have a fp32 executable of vulkan as of release v2.57.21

Use **cunet** model digital art unless another model is required for a specific purpose

Usage

DO NOT WAIFU UPSCALE DIGITAL RAWS UNLESS THERE ARE EXTREME CIRCUMSTANCES
DO NOT UNDER ***ANY*** CIRCUMSTANCES ATTEMPT TO DENOISE OR UPSCALE ARTIST PROVIDED TRUE PNG RAWS
Waifu2x denoise should only be used for denoising raws that ***actually*** have noise

- The models are trained for jpg noise
- Though not designed for it, it may also be able to achieve useful results on certain levels of webp noise

Likely denoise levels:

- 0: suitable for jpg quality 93+
- 1: suitable for jpg quality 75-90
- 2: suitable for jpg quality 75 and below
- 3: simply strong denoise, works on all jpg levels

This jpg quality level is a 1-100 value to represent the intensity of the rounding in the quantization step of jpg encoding
The levels provided are just estimates, different image output programs may have different quantization qualities for the same "quality" marker
Run comparison tests with hyper level afterwards to check if the denose level was actually suitable
often times, if you need level 2 it is not unreasonable just use level 3

quality 95+ may possibly be denoised effectively with methods other than waifu
Most common raws jpg quality levels are often 75-90

Other settings that affect denoise quality:
Split/Block/Tile size

- How much of the image is taken into processed at once (split size x split size square box)
- For cunet model, greater split size results in a better denoise result
- Higher value can also increase the processing speed
- 800 is around the limit for 4gb vram at fp32 - ncnn vulkan cannot(?) use more than 4gb per chunk
    - (on a rx580 8gb on win10. higher settings may or may not work on other system configurations)

TTA (Time Test Augmentation)

- Technical def: creating augmented copies for each item in the dataset with the goal of a more accurate prediction by returning the combined results
- The current implementation of the setting in waifu2x creates 8 copies of the image (original, rotate 90, 180, 270, and mirror for each of those) and returns the average of the result for the 8 passes

Floating point precision (fp32 or fp16)

- (on ncnn vulkan only) 16bit floating precision allows for faster speed on supported hardware and lower vram usage at the cost of artifacting resulting from floating point rounding
- Floating point 32 bit precision is the native for the original waifu2x code and models
    - Ncnn vulkan release does not ship with complete fp32 compiled executable
        - The gui addin contains a separate exe compiled with fp32 options
        - Or you can choose to build from source with the fp32 flags set

Using a denoise level that is not suitable for the input noise level will result in a worse denoise result

- Ex. if lvl 2 is used on a jpg 90, the results will be worse than the output of lvl 1
    - See sample images for example
- Denoise 3 is a special case, it is simply denoise strongly, and can be used for any level
    - Keep in mind that grainy artstyle/intentional noise filters/certain types of high frequency details may suffer from strongly denoising

After waifu denoise, ensure that the color profile of the output image is correct (sGray for grayscale, sRGB for color)
Check with hyper level for remnants of noise that still need to be removed and whether the denoise level was suitable

Notes:
A number of other details can be found on the waifu caffe english translated readme
Cpu mode on caffe

- Greater compatibility (runs on cpu), but higher memory cost and extremely slow
- Denoise output pretty much exactly equal to caffe gpu mode

Quality of ncnn vulkan

- Caffe is slightly quantitatively better, but at the proper settings ncnn vulkan is effectively the same/very close qualitatively, even under hyperlevel

Extras:

What is compression? Is it bad?

- Compression is the process of encoding information using fewer bits than the original representation.
    - It can be both lossy or lossless depending on implementation
- It is not a bad thing, in fact it is pretty much required with the amount of data that is stored and transferred across the net
    - The problem comes in when lossy (or sometimes even lossless) methods are used for applications where they are not suitable

What is a jpg? Why does it need denoising?

- JPG/JPEG is an image compression method
    - The common implementation of jpg is pretty much always lossy, meaning information is discarded for smaller file sizes
    - DCT on 8x8 blocks and rounding to integer (basic idea)
- Computerphile has a good basic rundown of jpg encoding; the 4th video explains the drawbacks when used for text and digital art
    - How JPEG Works
    - Jpg can have very useful applications for camera photos, but some of the assumptions about camera images made when designing the standard dont hold when applied to text and/or digital art
- It is ***possible*** for Jpegs to be re saved for a ***second*** time without ***further*** loss to data, but this must be done with extreme care to a variety of factors, as any changes made to the quantization table or any misalignment to the blocks will result in forced requantization for the entire image and result in stacking jpg losses
    - In addition, any pixels that are edited will need to be lossily encoded

What is png? Why is it recommended for digital art?

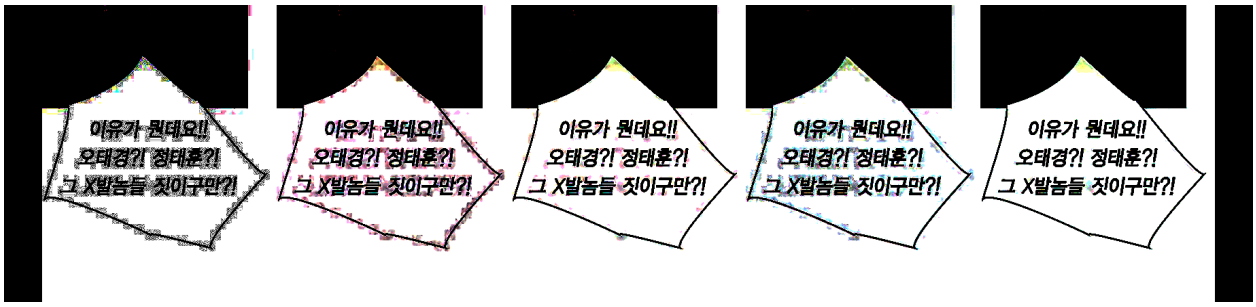
- It is lossless (when not force saved in palette mode) and supports transparency (though transparency is not usually a factor in manga/webtoons)

What about webp?

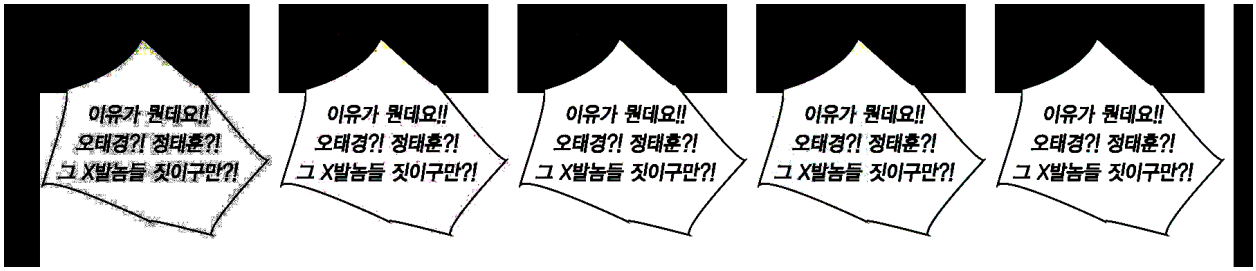
🍎 Support not widespread enough yet 🍎

Sample Images

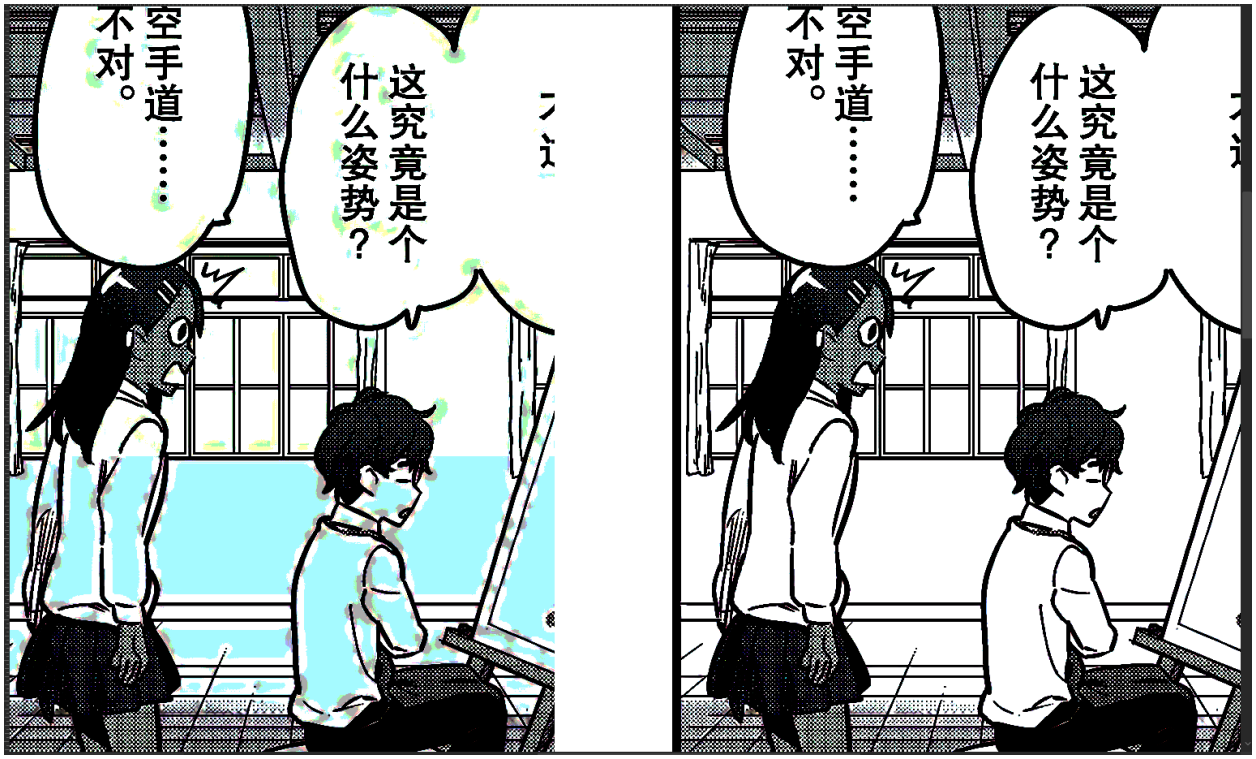
Example of denoise levels (original, 0, 1, 2, 3) on a jpg quality 90 hyper leveled below



The above sample leveled to 252 on highlights post denoise (still hyperleveled)



Example of fp16 (left) vs fp32 (right) hyper leveled below



Example of jpg noise (on white) hyper leveled below
The slider can be dragged to the other side to see the noise on blacks