EUROGRAPHICS 2023 / K. Myszkowski and M. Nießner
(Guest Editors)

Volume 42 (2023), Number 2

Subpixel Deblurring of Anti-Aliased Raster Clip-Art

J. Yang[1], N. Vining[1,2], S. Kheradmand[1], N. Carr[3], L. Sigal[1], and A. Sheffer[1]

[1]University of British Columbia, Canada
[2]NVIDIA, Canada
[3]Adobe, United States of America

Appendix A: Implementation Details

Dataset.

We assemble a dataset of 145 vector images by collecting vector clip-art from online repositories. These consist of a variety of complex shapes comprising single or multiple objects; different color palette sizes and different complexity (some with as few as 4 regions, others with over a hundred). We split this dataset into disjoint train/test subsets, with training composed of 68 and testing of 77 images.

We rasterize vector images at different $n \times n$ resolutions in [Ado17] using the supersampling anti-aliasing setting designed for artwork (we explore the font hinted setting in Appendix B as an ablation). Additionally, we rasterize each input at double resolution $2n \times 2n$ with no anti-aliasing. Following the cross-resolution consistency principle, on training data we use these double-resolution inputs as ground truth, and use them for quantitative evaluation for test data.

Data Augmentation. In order to augment the training data, we transform each image pair by rotations, reflections and switching of RGB color channels. As a result, each distinct image in our training dataset has 72 variations in training data. Each distinct test image also has 72 variations, which we use for denoising pix2pix outputs (as described in more detail in Section 4 of our paper).

Preprocessing. We add two rows of background colored pixels around each input prior to processing; we define the background color as the most common color along the image perimeter. This process makes the background a single region. We remove this padding from the final inputs. In our experiments adding padding improved the performance of both steps of our method.

Architecture Details.

We inherit architectural design and corresponding inductive biases from pix2pix. This includes U-Net architecture for the generator (see Fig. 8), that preserves pixel correspondence and locality, and a default patch discriminator (see Fig. 9) with an additional L1 loss. Our own inductive biases for architecture design focused on perceptual color space (LAB) for the loss function computation and gradient prediction (as opposed to direct prediction) of high resolution content (illustrated in both Figs. 8 and 9).

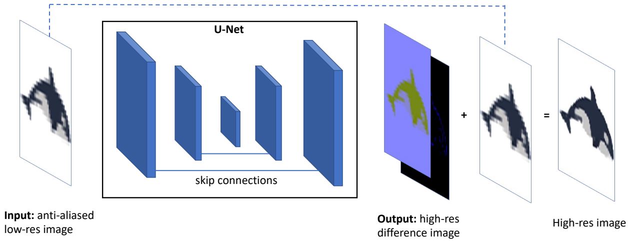
U-Net
skip connections
+
=
Input: anti-aliased low-res image
Output: high-res difference image
High-res image

Figure 1: *Our U-Net architecture adopted from [IZZE17].*

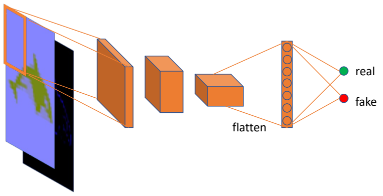

Figure 2: *Our patch discriminator adopted from [LW16].*

Training Details.

We train separate models for different resolutions setting n equal to $16px$, $32px$, $64px$ and $128px$. In order to upsample $n \times n$ anti-aliased images to the pix2pix input resolution of $2n \times 2n$ we use nearest-neighbour upsampling, copying one pixel in the $n \times n$ input image to four pixels in the $2n \times 2n$ counterpart. This simple upsampling is motivated by our desire to keep the original anti-aliasing and not to introduce additional interpolations into the input. If needed, such interpolations can be learned by the pix2pix network itself.

We use a ResNet backbone, with 6 residual blocks, for *pix2pix* itself (the predictor) and leverage PatchGAN [LW16] as a discriminator. We optimize networks for 300 epochs with batch size of 16 using the Adam optimizer [KB14]. We tune the learning rate for each resolution, as it is resolution dependent, and employ a learn-

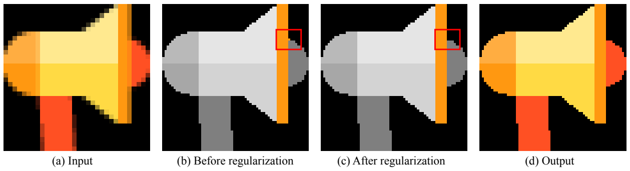

Figure 3: *Our regularization step removes redundant pixel-wide protrusions along region boundaries (b,c).*

ing schedule where this rate is fixed for the first 150 epochs and then linearly reduced to 0 over the remaining 150.

Color Space Distances

Measuring color-space distances in a manner consistent with human perception remains an open problem [SASS14]. In our colorization step we use a combination of established space metrics and heuristics based on observations of our training data. Specifically, unless stated otherwise, we use OkLab [Ott21] distance for all measurements. We overcome minute variations in pixel color by defining two colors as *the same* if the distance between them in RGB space is less than or equal to $||(2,2,2)||$. While in our experiments OkLab distances are generally well aligned with viewer perception for colors which are farther apart, we found them too sensitive for dark colors. Accordingly, if two colors both have RGB space values between $(0,0,0)$ and $(20,20,20)$ and the RGB space norm of their difference is below $||(20,20,20||$, we set the distance between them to zero.

We define pixels as outliers if their color is at least $\varepsilon = ||(5,5,5)||$ apart from the closest affine combination of its neighbor colors in RGB space.

Boundary Regularization

As noted in Sec. 5.2, our simplicity enforcement step removes non-simple regions but can undesirably elongate region boundaries, and can in particular introduce single-pixel-wide protrusions. Since viewers are unlikely to hallucinate such protrusions, we seek to remove them by merging them with a neighboring region. We identify protrusions which are one pixel wide and two or more pixels long, ignoring ones which are part of constant slope lines. We merge the protrusion with the neighboring region of the most similar color if doing so does not introduce longer protrusions.

16×16 Inputs

Extremely low resolutions pose unique challenges, both for learning low-blur magnifications and for detecting patch seeds. The first challenge arises since the pix2pix network uses a kernel size of 3x3 and a fixed number of kernels per residual block. As an artifact the receptive field is fairly large and with extremely low resolution inputs the convolutional nature of the operations is effectively lost. We address this challenge when training our network on $16px$ data by first magnifying our inputs using nearest neighbor sampling to $32px$ and our outputs to $64px$ accordingly. At run-time, after running our approach we then sub-sample the $64px$ outputs back to $32px$ to produce the final result by using the median of each block of 4 neighboring pixels. Aside from this input/output magnification, we used the same data augmentation process, and train the network with the same hyperparameters as for other resolutions.

Using our default patch seed detection on $16px$ inputs is similarly problematic, as the number of pixels occupied by original regions drops dramatically (Fig. **??**, top); keeping our default criterion leads to a loss of information encoded in long one-pixel wide regions (e.g. the princess's eyes in Fig. **??**). Accordingly, for $16px$ inputs we redefine the patch seeds to include all pairs of adjacent same-color pixels. The rest of the processing remains the same.

Runtimes

Our training times are resolution dependent. Training the $16px$ and $32px$ networks took around 2.5 hours; training the $64px$ network took around 6 hours, and training the $128px$ network took around 24 hours. Our models were trained on a GeForce RTX 2080.

Our method's run-time is dominated by the coloring step (Sec. **??**). Our median run-times are 0.6 seconds for $16px$ inputs, 3.5 seconds for $32px$ inputs, 33 seconds for $64px$ inputs, and 6.8 minutes for $128px$ inputs. Timings were measured on a Intel Core I7-8700k running at 3.70GHz with 32GB of system memory.

Appendix B: Ablations

Invariance to Rasterization For the experiments in the paper so far, we used inputs rasterized using standard supersampling based anti-aliasing; supersampling is the default method used for rasterizing clip-art images [FVVD*96]. Our blur-free magnification technique does not, however, assume any specific rasterization scheme and can adapt to differences in rasterization within pix2pix learning. To illustrate this, we also conduct experiments with font hinting as the rasterization mode. The results can be seen in Figure 11 and were produced with no *pix2pix* retraining. Despite the clear differences in the inputs induced by the two different rasterization techniques, our approach successfully produces anti-aliasing free outputs that are sharp and structurally consistent in both cases.

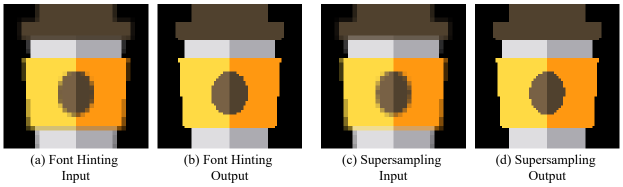

Figure 4: *Results on inputs rasterized using different schemes: (a) input produced using supersampling based anti-aliasing (b) input produced using font hinting based anti-aliasing (c) output for (a); (d) output for (b).*

Colorization Energy Our colorization energy $E(I_o)$ combines four terms measuring color distinction E_d, compactness E_C, cross-resolution consistency E_a, and seed anchoring E_s; we use the weights $w_a = 10$ and $w_b = 0.5$ to balance these terms. In our experiment (Fig. 12) we increased or decreased each of these weights

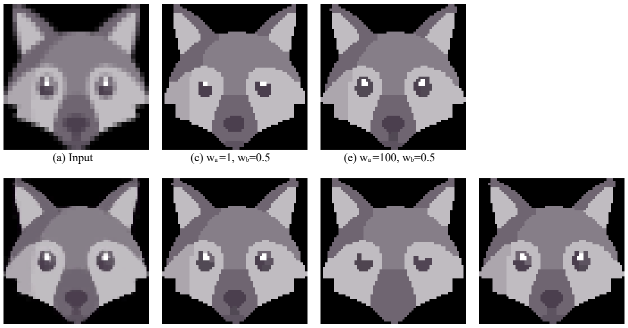
(a) Input
(c) w_a=1, w_b=0.5
(e) w_a=100, w_b=0.5
(b) low-blur magnification
(d) w_a=10, w_b=0.05
(f) w_a=10, w_b=5
(g) Our Method

Figure 5: *Results of increasing and decreasing the weights w_a, w_b by a factor 10.*

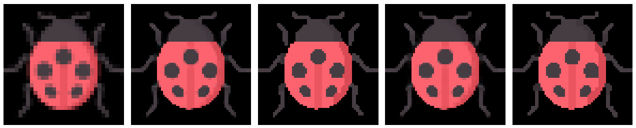
(a) Input
(b) learned low-blur magnification in RGB space
(c) Output using (b)
(d) learned low-blur magnification in LAB space
(e) Our Output

Figure 6: *Color space impact: (left) RGB space magnification; (right) OkLab pace magnification; while our method recovers the top part of the vertical line on ladybug's back, it gets removed by the RGB space method.*

by a factor of 10. Decreasing w_a or increasing w_c increases the importance of the compactness term, decreasing the number of output regions, while the inverse changes results in the preservation of redundant details. Our output balances the conflicting cues in a manner consistent with viewer expectations.

Impact of Color Space Our pix2pix network is trained using the LAB color space [FVVD*96]. Fig. 13 compares our results to those produced using a network trained in RGB space. The differences, while minor overall, can impact the recovery of fine details when the color difference between fine details and adjacent regions is not very large.

Comparison vs Real-ESRGAN. Fig. 14 shows the impact of replacing our first step, based on pix2pix, with nearest-neighbour upsampling and then deblurring based on the Real-ESRGAN model, retrained on our inputs. As the image shows, the blur and large color variation in their outputs means that we are no longer to reliably detect outliers and patches in the outputs of the learning step. Consequently, our palette computation is unable to extract a meaningful palette from this data.

Appendix C: Study Setup

We detail below the protocols used for the three studies reported on in the paper. All study data is provided in the supplementary.

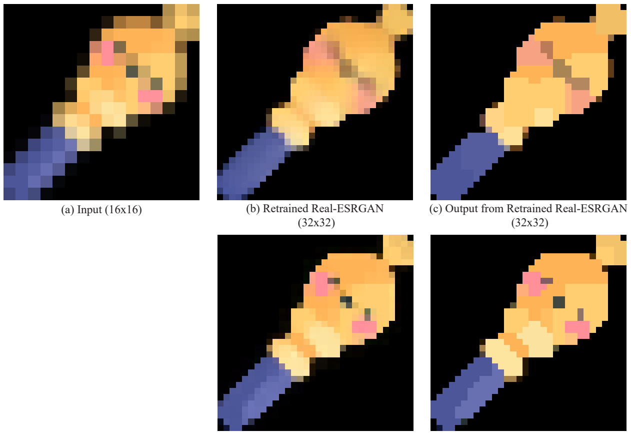
(a) Input (16x16)
(b) Retrained Real-ESRGAN (32x32)
(c) Output from Retrained Real-ESRGAN (32x32)
(d) Our Intermediate Solution (32x32)
(e) Our Output (32x32)

Figure 7: *Replacing our pix2pix network with Real-ESRGAN retrained on our inputs.*

Perception of Anti-Aliased Images

Color Palette Study

Our first informal study aimed to understand how human observers perceive the size of the artist-intended color palettes in anti-aliased clip-art. Participants in this study were presented with 10 anti-aliased, low resolution images and were asked the question "Mentally remove the anti-aliasing blur. How many distinct colors does the deblurred image have?" They were presented with two basic examples (two diagonally placed rectangles, with three colors total in the image; and one single color "O" shape with two colors total in the image); no other instructions were provided. The study included six participants, 3 male and 3 female.

In all cases, participants perceived input images as having small palette sizes, with answers that were largely consistent across all inputs and closely matching the number of colors in the originating vector images. When participants did not correctly identify the number of colors in the originating vector image, they tended to slightly underestimate, rather than overestimate, the number of colors used. This study confirms our focus on compact color palettes as a key property of the mental images viewers conjure when presented with anti-aliased clip-art. The survey and participant answers are included in the supplementary material.

Segmentation Study

Our second informal study aimed to understand how human observers mentally segment anti-aliased clip-art images. (Fig. 2, in paper.) Participants in this study were presented with 6 images and were asked to ”Mentally deblur and magnify this image. Trace the outlines of the single color regions in the blur-free output you envisioned. Please pay attention to details.” They were presented with two basic tracing examples (single color ”O” shapes and two diagonally placed rectangles); no other instructions were provided. The study included five participants, four male and one female.

Participants' traced outputs were largely consistent, with some variation in details, and were largely closely aligned with the region boundaries in the blur-free double resolution rasterizations of the underlying inputs. Participants did not hallucinate regions that were not evident in the input. The outputs were therefore consistent with our hypothesis of cross-resolution consistency and simplicity as major factors in perception of anti-aliased clip art imagery. Our outputs on the inputs traced by the participants are included in the supplementary, and are well aligned with the manual tracing outputs.

Comparative Study

In our comparative study, participants were shown input images, together with our result and an alternative result using the following layout. The input was shown at the top and marked as ‘A’, and the two magnified outputs were placed at the bottom and marked as ‘B’ and ‘C’. The order of the magnified outputs on the bottom was randomized. Participants were then asked to “Mentally deblur and magnify the anti-aliased raster image on the top (A). Which of the images on the bottom (B or C) comes closest to the blur-free image you mentally assembled? Please zoom in to see the differences.” The possible answer options were “B”, “C”, “Both”, and “Neither”. They were shown two ground truth examples: in one option, participants were shown the ground truth output and an anti-aliased double resolution rasterization of the originating image; in the other they were shown the ground truth image and a nearest neighbor magnification of the input. Participants were shown the answers to those. For VectorMagic we used the setting of “artwork with blended edges”, “high quality” and “unlimited color’ which is recommended for anti-aliased clip art and which produced the best results. We used default parameters for all other methods. The study included 70 participants, 53 male and 17 female. The complete list of questions and answer breakdowns are included in the supplementary.

References

[ABA*16] ARTUSI A., BANTERLE F., AYDIN T., PANOZZO D., SORKINE-HORNUNG O.: *Image Content Retargeting: Maintaining Color, Tone, and Spatial Consistency*. CRC Press, 2016.

[Ado17] ADOBE: Adobe Illustrator 2017: Image Trace. http://www.adobe.com/, 2017. 1, 5

[Ado21] ADOBE: Adobe Photoshop Lightroom. Super Resolution. https://www.adobe.com/products/photoshop-lightroom/super-resolution.html, 2021.

[BK04] BOYKOV Y., KOLMOGOROV V.: An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. *IEEE PAMI 26*, 9 (2004), 1124–1137.

[Die08] DIEBEL J. R.: *Bayesian image vectorization: The probabilistic inversion of vector image rasterization*. Ph.D. dissertation, Stanford Univ., 2008.

[DLHT15] DONG C., LOY C. C., HE K., TANG X.: Image super-resolution using deep convolutional networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2 (2015), 295–307.

[DSG*20] DOMINICI E., SCHERTLER N., GRIFFIN J., HOSHYARI S., SIGAL L., SHEFFER A.: Polyfit: Perception-aligned vectorization of raster clip-art via intermediate polygonal fitting. *ACM Transaction on Graphics 39* (2020).

[FH04] FELZENSZWALB P. F., HUTTENLOCHER D. P.: Efficient graph-based image segmentation. *International Journal of Computer Vision 59*, 2 (2004), 167–181.

[FLB16] FAVREAU J.-D., LAFARGE F., BOUSSEAU A.: Fidelity vs. simplicity: a global approach to line drawing vectorization. *ACM SIGGRAPH* (2016).

[FSH*06] FERGUS R., SINGH B., HERTZMANN A., ROWEIS S. T., FREEMAN W. T.: Removing camera shake from a single photograph. *ACM TOG 25*, 3 (2006), 787–794.

[FVVD*96] FOLEY J. D., VAN F. D., VAN DAM A., FEINER S. K., HUGHES J. F., HUGHES J.: *Computer graphics: principles and practice*, vol. 12110. Addison-Wesley Professional, 1996. 2, 3, 7

[HDS*18] HOSHYARI S., DOMINICI E., SHEFFER A., CARR N., CEYLAN D., WANG Z., SHEN I.-C.: Perception-driven semi-structured boundary vectorization. *ACM Transaction on Graphics 37*, 4 (2018). doi:10.1145/3197517.3201312.

[HEK21] HETTINGA G. J., ECHEVARRIA J., KOSINKA J.: Efficient Image Vectorisation Using Mesh Colours. In *Smart Tools and Apps for Graphics - Eurographics Italian Chapter Conference* (2021), Frosini P., Giorgi D., Melzi S., Rodolà E., (Eds.), The Eurographics Association. doi:10.2312/stag.20211484.

[Hyl11] HYLLIAN: Xbr. https://github.com/Hyllian/glsl-shaders/blob/master/xbr/shaders/xbr-lv2.glsl, 2011.

[Ink20] INKSCAPE: Inkscape, 2020. URL: https://inkscape.org.

[IZZE17] ISOLA P., ZHU J.-Y., ZHOU T., EFROS A. A.: Image-to-image translation with conditional adversarial networks. *CVPR* (2017). 1, 6

[KB14] KINGMA D. P., BA J.: Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014). 1, 6

[KHEK76] KUWAHARA M., HACHIMURA K., EIHO S., KINOSHITA M.: Processing of ri-angiocardiographic images. In *Digital processing of biomedical images*. Springer, 1976, pp. 187–202.

[KKD09] KYPRIANIDIS J. E., KANG H., DÖLLNER J.: Image and video abstraction by anisotropic kuwahara filtering. *Computer Graphics Forum 28*, 7 (2009), 1955–1963. Special issue on Pacific Graphics 2009. doi:10.1111/j.1467-8659.2009.01574.x.

[KKT20] KIM W., KANEZAKI A., TANAKA M.: Unsupervised learning of image segmentation based on differentiable feature clustering. *IEEE Transactions on Image Processing* (2020).

[KL11] KOPF J., LISCHINSKI D.: Depixelizing pixel art. *ACM TOG 30*, 4 (2011), 99:1–99:8.

[Kof55] KOFFKA K.: *Principles of Gestalt Psychology*. International library of psychology, philosophy, and scientific method. Routledge & K. Paul, 1955.

[LCS*21] LIANG J., CAO J., SUN G., ZHANG K., GOOL L. V., TIMOFTE R.: SwinIR: Image restoration using swin transformer, 2021. arXiv:2108.10257.

[LL06] LECOT G., LEVY B.: Ardeco: Automatic Region Detection and Conversion. In *EGSR* (2006), pp. 349–360.

[LSZ*21] LIANG J., SUN G., ZHANG K., VAN GOOL L., TIMOFTE R.: Mutual affine network for spatially variant kernel estimation in blind image super-resolution. In *IEEE International Conference on Computer Vision* (2021).

[LTH*17] LEDIG C., THEIS L., HUSZAR F., CABALLERO J., CUNNINGHAM A., ACOSTA A., AITKEN A., TEJANI A., TOTZ J., WANG Z., SHI W.: Photo-realistic single image super-resolution using a generative adversarial network. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)* (July 2017).

[LW16] LI C., WAND M.: Precomputed real-time texture synthesis with markovian generative adversarial networks. *ECCV* (2016). 1, 6

[LWDF09] LEVIN A., WEISS Y., DURAND F., FREEMAN W. T.: Understanding and evaluating blind deconvolution algorithms. *CVPR* (2009).

[MG21] MCGUIRE M., GAGIU M.: MMPX style-preserving pixel art magnification. *Journal of Graphics Techniques* (January 2021), 36. Journal of Graphics Techniques.

[MRC*20] MA C., RAO Y., CHENG Y., CHEN C., LU J., ZHOU J.: Structure-preserving super resolution with gradient guidance, 2020. arXiv:2003.13081.

[OBW*08] ORZAN A., BOUSSEAU A., WINNEMÖLLER H., BARLA P., THOLLOT J., SALESIN D.: Diffusion curves: A vector representation for smooth-shaded images. *ACM TOG 27*, 3 (2008).

[Ott21] OTTOSSON: Oklab. https://bottosson.github.io/posts/oklab/, 2021. 2, 6

[RFB15] RONNEBERGER O., FISCHER P., BROX T.: U-net: Convolutional networks for biomedical image segmentation. *MICCAI* (2015).

[RGLM21] REDDY P., GHARBI M., LUKAC M., MITRA N. J.: Im2Vec: Synthesizing vector graphics without vector supervision. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (Los Alamitos, CA, USA, jun 2021), IEEE Computer Society, pp. 7338–7347. URL: https://doi.ieeecomputersociety.org/10.1109/CVPR46437.2021.00726, doi:10.1109/CVPR46437.2021.00726.

[SASS14] STONE M., ALBERS SZAFIR D., SETLUR V.: An engineering model for color discriminability as a function of size. In *IS&T 22nd Color Imaging Conference* (2014). 2, 6

[SBv05] SÝKORA D., BURIÁNEK J., ŽÁRA J.: Sketching cartoons by example. In *Proc. Sketch-Based Interfaces and Modeling* (2005), pp. 27–34.

[SLWS07] SUN J., LIANG L., WEN F., SHUM H.-Y.: Image vectorization using optimized gradient meshes. In *ACM SIGGRAPH* (2007).

[Ste03] STEPIN M.: HQX. http://web.archive.org/web/20070717064839/www.hiend3d.com/hq4x.html, 2003.

[Sto03] STONE M. C.: *A Field Guide to Digital Color*. CRC Press, 2003.

[TFCRS11] THOMPSON W., FLEMING R., CREEM-REGEHR S., STEFANUCCI J. K.: *Visual Perception from a Computer Graphics Perspective*, 1st ed. A. K. Peters, Ltd., USA, 2011.

[TM98] TOMASI C., MANDUCHI R.: Bilateral filtering for gray and color images. In *Sixth International Conference on Computer Vision (IEEE Cat. No.98CH36271)* (1998), pp. 839–846. doi:10.1109/ICCV.1998.710815.

[Vec17] VECTOR MAGIC:. Cedar Lake Ventures http://vectormagic.com/, 2017.

[VPB*22] VINKER Y., PAJOUHESHGAR E., BO J. Y., BACHMANN R. C., BERMANO A. H., COHEN-OR D., ZAMIR A., SHAMIR A.: Clipasso: Semantically-aware object sketching. *ACM Trans. Graph. 41*, 4 (2022).

[WEK*12] WAGEMANS J., ELDER J. H., KUBOVY M., PALMER S. E., PETERSON M. A., SINGH M., VON DER HEYDT R.: A century of gestalt psychology in visual perception i. perceptual grouping and figure-ground organization. *Psychological Bulletin 138*, 6 (2012), 1172–1217.

[WXDS21] WANG X., XIE L., DONG C., SHAN Y.: Real-ESRGAN: Training real-world blind super-resolution with pure synthetic data. In *International Conference on Computer Vision Workshops (ICCVW)* (2021).

[WZGW17] WANG C., ZHU J., GUO Y., WANG W.: Video vectorization via tetrahedral remeshing. *IEEE TIP 26*, 4 (April 2017), 1833–1844.

[XK17] XIA X., KULIS B.: W-net: A deep model for fully unsupervised image segmentation. *arXiv: 1711.08506* (2017).

[XLXJ11] XU L., LU C., XU Y., JIA J.: Image smoothing via l 0 gradient minimization. In *Proceedings of the 2011 SIGGRAPH Asia conference* (2011), pp. 1–12.

[XLY09] XIA T., LIAO B., YU Y.: Patch-based image vectorization with automatic curvilinear feature alignment. *ACM TOG 28*, 5 (2009).

[XSTN14] XIE G., SUN X., TONG X., NOWROUZEZAHRAI D.: Hierarchical diffusion curves for accurate automatic image vectorization. *ACM Trans. Graph. 33*, 6 (2014), 230:1–230:11.

[YCZ*16] YANG M., CHAO H., ZHANG C., GUO J., YUAN L., SUN J.: Effective clipart image vectorization through direct optimization of bezigons. *IEEE TVCG 22*, 2 (2016), 1063–1075.

[ZCZ*09] ZHANG S.-H., CHEN T., ZHANG Y.-F., HU S.-M., MARTIN R. R.: Vectorizing cartoon animations. *IEEE TVCG 15*, 4 (2009), 618–629.

[ZLVGT21] ZHANG K., LIANG J., VAN GOOL L., TIMOFTE R.: Designing a practical degradation model for deep blind image super-resolution. In *IEEE International Conference on Computer Vision* (2021).