Improving the Perceptual Quality of 2D Animation Interpolation

Shuhong Chen, Matthias Zwicker
{shuhong, zwicker}@cs.umd.edu

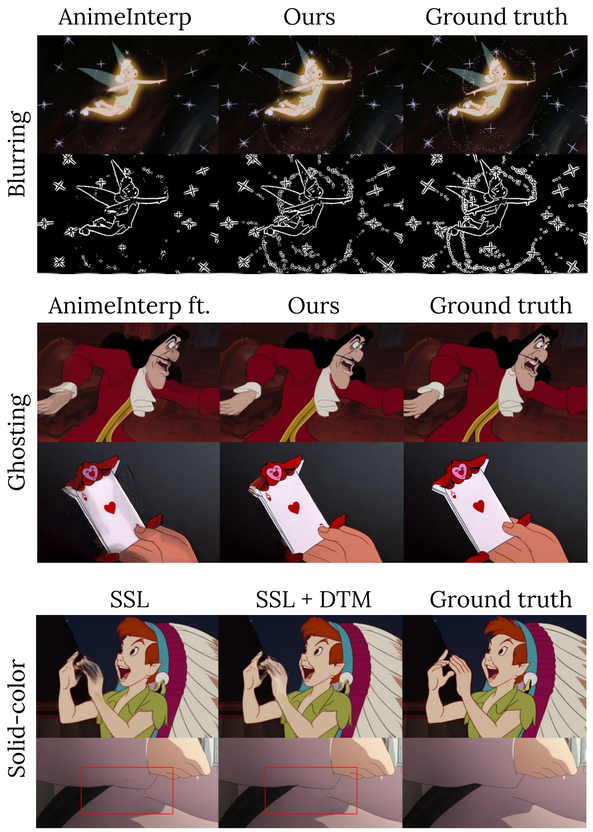
UNIVERSITY OF MARYLAND

European Conference on Computer Vision, ECCV 2022

Abstract
Traditional 2D animation is labor-intensive, often requiring animators to manually draw twelve illustrations per second of movement. While automatic frame interpolation may ease this burden, 2D animation poses additional difficulties compared to photorealistic video. In this work, we address challenges unexplored in previous animation interpolation systems, with a focus on improving perceptual quality. Firstly, we propose SoftsplatLite (SSL), a forward-warping interpolation architecture with fewer trainable parameters and better perceptual performance. Secondly, we design a Distance Transform Module (DTM) that leverages line proximity cues to correct aberrations in difficult solid-color regions. Thirdly, we define a Restricted Relative Linear Discrepancy metric (RRLD) to automate the previously manual training data collection process. Lastly, we explore evaluation of 2D animation generation through a user study, and establish that the LPIPS perceptual metric and chamfer line distance (CD) are more appropriate measures of quality than PSNR and SSIM used in prior art.

Challenges from prior art
- Perceptual artifacts (blurring, ghosting) →
- Convnets unsuited to textureless cel-coloring →
- Requires expensive manual data collection →
- Evaluation unrepresentative of animation quality →

Our approach
(A) <u>*SoftsplatLite*</u> architecture reduces artifacts
(B) <u>*Distance Transform Module*</u> refines solid-color regions
(C) <u>*Restricted Relative Linear Discrepancy*</u> automates data collection
(D) <u>*User study*</u> establishes LPIPS & Chamfer Dist. as quality metrics



(A) <u>SoftsplatLite</u> reduces blurring and ghosting by redesigning the synthesis network.

While prior work like AnimeInterp focused on improving optical flow for 2D animation, design flaws in the synthesis network led to perceptual artifacts like blurring and ghosting. Our proposed SSL architecture combats these with improvements including LPIPS training, occlusion-mask infilling, and feature extraction pretraining. These modifications result in better perceptual performance as well as 35% less trainable parameters than AnimeInterp.

Occlusion-mask infilling

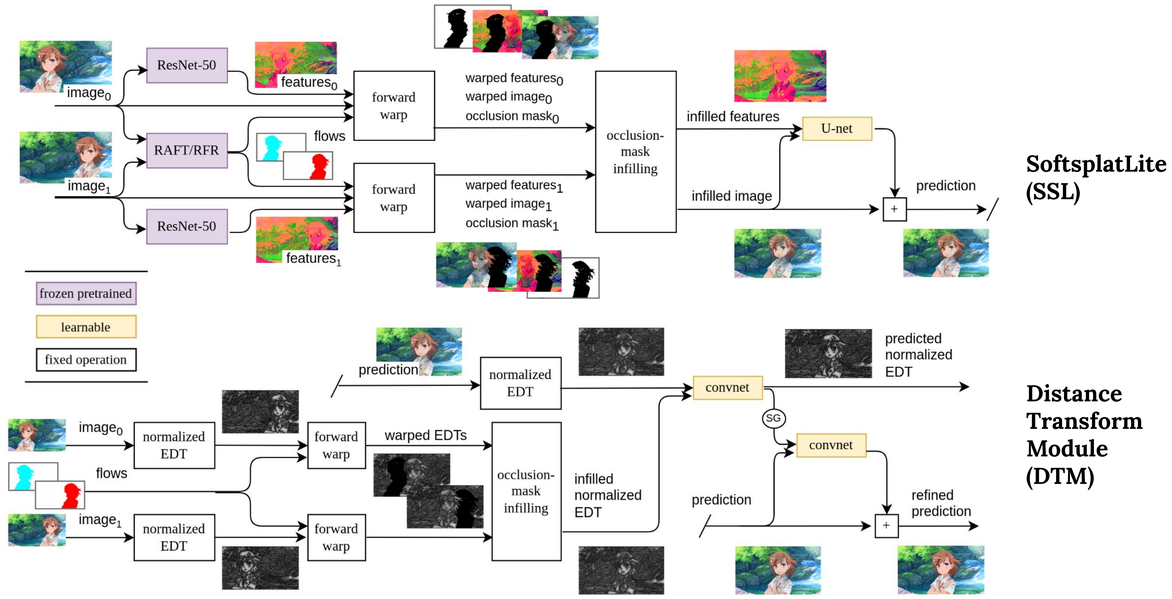
$$F = \frac{1}{2}\left(M_{0\to t}W_{0\to t}(f(I_0)) + (1 - M_{0\to t})W_{1\to t}(f(I_1))\right) + \frac{1}{2}\left(M_{1\to t}W_{1\to t}(f(I_1)) + (1 - M_{1\to t})W_{0\to t}(f(I_0))\right)$$

(B) <u>Distance Transform Module</u> refines solid-color regions by leveraging line proximity cues.

Cel-colored frames have more solid-color regions than natural images, posing a challenge to convnets that favor texture. To improve performance in these areas, we design a refinement module that leverages auxiliary information from the normalized euclidean distance transform (NEDT) of a frame's extracted diff-of-gaussians (DoG) line drawing.

$$DoG(I) = \frac{1}{2} + t(G_{k\sigma}(I) - G_{\sigma}(I)) - \epsilon,$$

$$NEDT(I) = 1 - \exp\{\frac{-EDT(DoG(I) > 0.5)}{\tau d}\},$$



Linear & evenly-spaced ⇒ low RRLD ✓

Non-linear or unevenly-spaced ⇒ high RRLD ✗

Criteria	Prefer Ours	Lower LPIPS	Lower CD	Higher PSNR	Higher SSIM
cleaner/sharper lines	86.01%	86.56%	78.20%	18.95%	15.48%
more consistent shape/color	78.82%	79.26%	73.99%	25.02%	22.66%
better overall quality	81.11%	81.55%	75.67%	22.97%	19.88%

Model	All				Eastern		Western	
	LPIPS	CD	PSNR	SSIM	LPIPS	CD	LPIPS	CD
DAIN	4.695	5.288	28.840	95.28	5.499	6.537	4.204	4.524
DAIN ft.	4.137	4.851	29.040	95.27	4.734	5.888	3.771	4.217
RIFE	4.451	5.488	28.515	95.14	4.933	6.618	4.156	4.796
RIFE ft.	4.233	5.411	27.977	93.70	4.788	6.643	3.894	4.658
ABME	5.731	7.244	29.177	*95.54*	7.000	10.010	4.955	5.552
ABME ft.	4.208	4.981	29.060	95.19	4.987	6.092	3.732	4.302
AnimeInterp	5.059	5.564	<u>29.675</u>	<u>95.84</u>	5.824	7.017	4.590	4.674
AnimeInterp ft.	*3.757*	*4.513*	28.962	95.02	*4.113*	*5.286*	*3.540*	*4.039*
Ours	<u>3.494</u>	<u>4.350</u>	*29.293*	95.15	<u>3.826</u>	<u>4.979</u>	<u>3.291</u>	<u>3.966</u>

(C) <u>Restricted Relative Linear Discrepancy</u> automates data collection by robustly measuring non-linearity within frame triplets.

Unlike 30-60fps natural video, 2D animation is usually drawn at 12fps. This greatly increases non-linearity and uneven spacing between frames, disqualifying most animation data for interpolation training. While AnimeInterp manually collects an animation dataset usable for training, we propose RRLD to automatically collect data by programmatically measuring non-linearity and uneven spacing.

$$RRLD(\omega_{0\to t}, \omega_{1\to t}) = \frac{1}{|\Omega|}\sum_{(i,j)\in\Omega}\frac{||\omega_{0\to t}[i,j] + \omega_{1\to t}[i,j]||/2}{||\omega_{0\to t}[i,j] - \omega_{1\to t}[i,j]||},$$

(D) <u>Our perceptual user study</u> is the first to establish both LPIPS and Chamfer Distance as more appropriate metrics for 2D animation quality, rather than PSNR/SSIM.

While AnimeInterp focused on PSNR/SSIM evaluation, we found that in actuality users consistently preferred our animations with better LPIPS and Chamfer line distance than PSNR/SSIM. This validates our state-of-the-art perceptual performance on 2D animation interpolation. We hope our study steers future research towards perceptual performance for this artistic domain.

$$CD(X_0, X_1) = \frac{1}{2HWD}\sum X_0 DT(X_1) + X_1 DT(X_0)$$

Code available at:
`github.com/ShuhongChen/eisai-anime-interpolator`

References:
Siyao, L., Zhao, S., Yu, W., Sun, W., Metaxas, D., Loy, C. C., & Liu, Z. (2021). Deep animation video interpolation in the wild. In Proceedings of the IEEE/CVF CVPR (pp. 6587-6595).
Niklaus, S., & Liu, F. (2020). Softmax splatting for video frame interpolation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 5437-5446).

Artist attributions:
Left side: animations taken from ATD12k dataset provided in AnimeInterp (Li et al.)
Right side: Anohana (2011, A-1 Pictures), Konobi (2016, feel.)
Schematic: foreground (hariken: https://danbooru.donmai.us/posts/5378938), background (k.k.: https://danbooru.donmai.us/posts/789765)

Acknowledgements: The authors would like to thank Lillian Huang and Saeed Hadadan for their discussion and feedback, as well as NVIDIA for GPU support.