# Super SloMo: High Quality Estimation of Multiple Intermediate Frames for Video Interpolation

Huaizu Jiang, Deqing Sun, Varun Jampani
Ming-Hsuan Yang, Erik Learned-Miller, Jan Kautz

Presented by Aleksei Kalinov

04 October 2018

# Talk Outline

1. Preliminary Topics
   a. Motion Estimation
   b. Image Warping
   c. U-Net
2. Super SlowMo
   a. Solution Approach
   b. Architecture Description
   c. Results

# Talk Outline

1. Preliminary Topics
   a. Motion Estimation
   b. Image Warping
   c. U-Net
2. Super SlowMo
   a. Solution Approach
   b. Architecture Description
   c. Results

# Motion Estimation

**Optical Flow.** A pattern of apparent motion of objects, surfaces, and edges in a visual scene caused by the relative motion between an observer and a scene.

# Block Matching

For each patch in the reference frame we optimize matching error with target frame image patch w.r.t. coordinates of the latter.

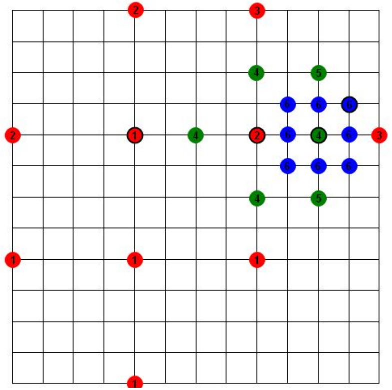
Reference frame

Target frame

# Block Matching Improvements

1. Hierarchical matching -- optimize at different resolutions
2. Sub-pixel matching -- better quality for fast-paced events.

3. Search region limit -- speed-up by introducing more assumptions
4. 2D Logarithmic search -- speed-up at a cost of quality
5. Millions of other searches -- area of research



# Differential methods

Express change in frames as finite difference equations. Solve for unknows.

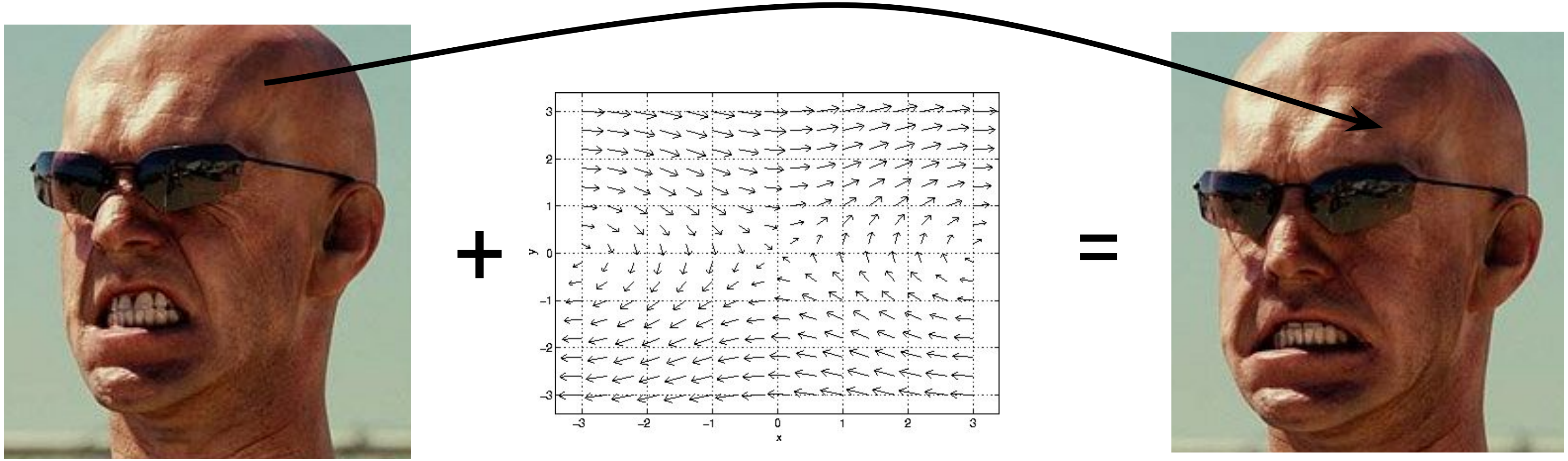
$$I_x(q_1)V_x + I_y(q_1)V_y = -I_t(q_1)$$

# Talk Outline

1. Preliminary Topics
    a. Motion Estimation
    b. Image Warping
    c. U-Net
2. Super SlowMo
    a. Solution Approach
    b. Architecture Description
    c. Results

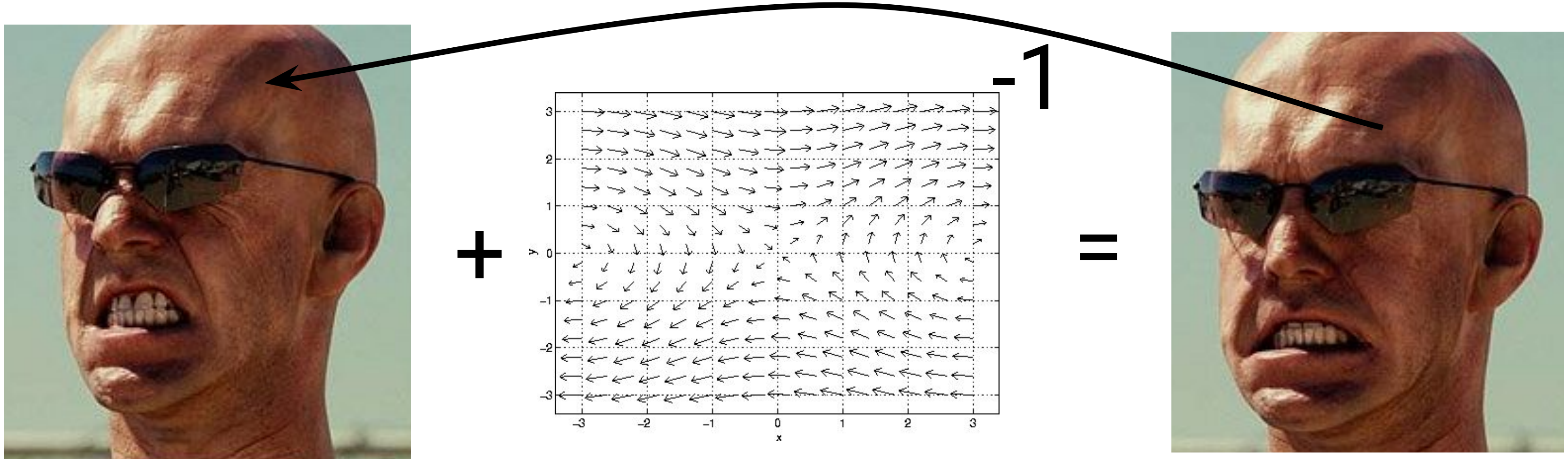
# Forward warping

Flow is used to warp pixel in the reference picture. Target location is rounded.



# Backward warping

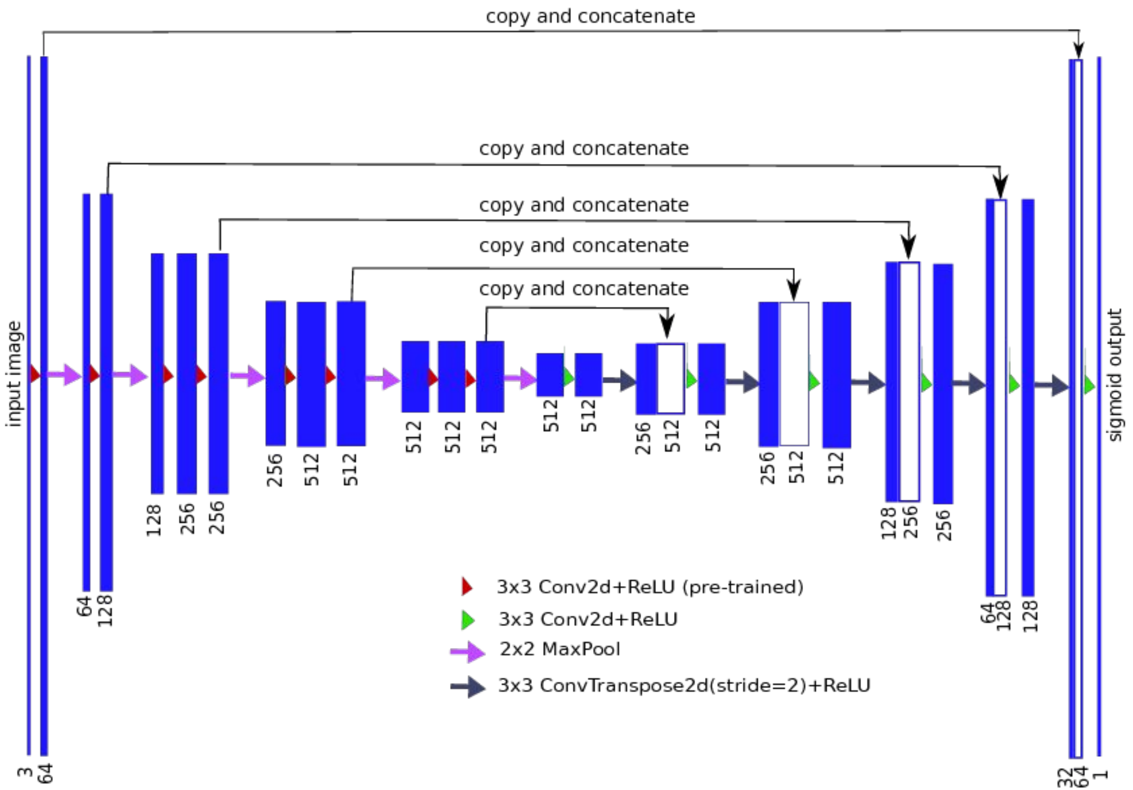
Each target pixel finds its original position using inverse flow. Intensity is interpolated.



# Talk Outline

1. Preliminary Topics
   a. Motion Estimation
   b. Image Warping
   c. U-Net
2. Super SlowMo
   a. Solution Approach
   b. Architecture Description
   c. Results

# U-Net


copy and concatenate
copy and concatenate
copy and concatenate
copy and concatenate
copy and concatenate
input image
sigmoid output
3
64
64
128
128
256
256
256
512
512
512
512
512
512
512
256
512
512
256
512
512
128
256
256
64
128
128
32
64
1
3x3 Conv2d+ReLU (pre-trained)
3x3 Conv2d+ReLU
2x2 MaxPool
3x3 ConvTranspose2d(stride=2)+ReLU

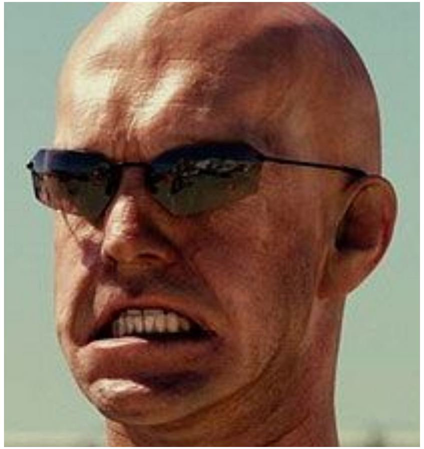
# Talk Outline

1. Preliminary Topics
    a. Motion Estimation
    b. Image Warping
    c. U-Net
2. Super SlowMo
    a. Solution Approach
    b. Architecture Description
    c. Results

# Problem statement

Given two images $I_0$ and $I_1$ and time $t$, predict and intermediate frame $I_t$.

?



T = 0

T = t

T = 1

# Solution

$$\hat{I}_t = \frac{1}{Z} \odot \left((1-t)V_{t\leftarrow 0} \odot g(I_0, F_{t\rightarrow 0}) + tV_{t\leftarrow 1} \odot g(I_1, F_{t\rightarrow 1})\right)$$

Warp

Warp

Occlude

Occlude

Linearly Combine

Go home

# What about flows?

Estimate flows for given images and interpolate for intermediate one.

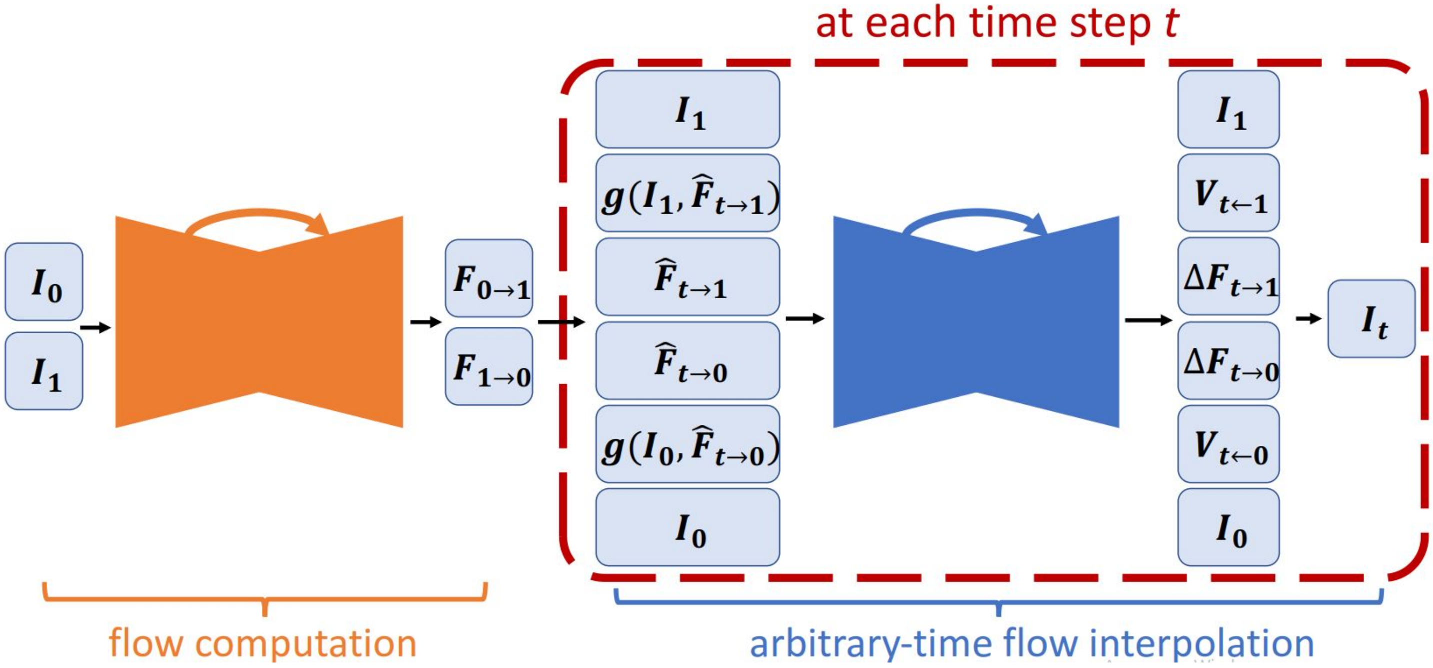
$$\hat{F}_{t\to 0} = -(1-t)tF_{0\to 1} + t^2 F_{1\to 0}$$

$$\hat{F}_{t\to 1} = (1-t)^2 F_{0\to 1} - t(1-t)F_{1\to 0}$$

# Talk Outline

1. Preliminary Topics
    a. Motion Estimation
    b. Image Warping
    c. U-Net
2. Super SlowMo
    a. Solution Approach
    b. Architecture Description
    c. Results

# Model architecture



# Flow interpolation network

Second network helps to resolve artifacts from quick motion.

# Training

Loss consists of 4 parts:

1. Reconstruction loss.
2. Perceptual loss.
3. Warping loss.
4. Smoothness loss.

Trained on ~300k Adobe and Youtube 240 fps videos.

$$l_r = \frac{1}{N}\sum_{i=1}^{N} \|\hat{I}_{t_i} - I_{t_i}\|_1$$

$$l_p = \frac{1}{N}\sum_{i=1}^{N} \|\phi(\hat{I}_t) - \phi(I_t)\|_2$$

$$l_w = \|I_0 - g(I_1, F_{0\to 1})\|_1 + \|I_1 - g(I_0, F_{1\to 0})\|_1 + \frac{1}{N}\sum_{i=1}^{N} \|I_{t_i} - g(I_0, \hat{F}_{t_i\to 0})\|_1 + \frac{1}{N}\sum_{i=1}^{N} \|I_{t_i} - g(I_1, \hat{F}_{t_i\to 1})\|_1$$

$$l_s = \|\nabla F_{0\to 1}\|_1 + \|\nabla F_{1\to 0}\|_1$$

# Talk Outline

1. Preliminary Topics
    a. Motion Estimation
    b. Image Warping
    c. U-Net
2. Super SlowMo
    a. Solution Approach
    b. Architecture Description
    c. Results

# Results in numbers

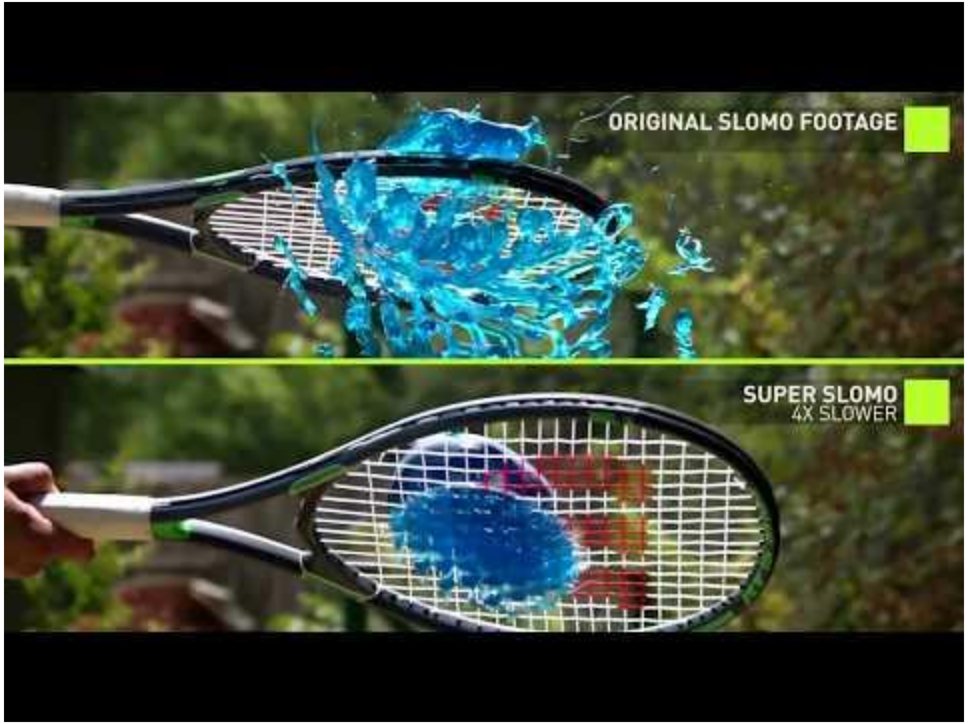
| | PSNR | SSIM | IE |
|---|---|---|---|
| Phase-Based [18] | 32.35 | 0.924 | 8.84 |
| FlowNet2 [1, 9] | 32.30 | 0.930 | 8.40 |
| DVF [15] | 32.46 | 0.930 | 8.27 |
| SepConv [20] | 33.02 | 0.935 | 8.03 |
| Ours (Adobe240-fps) | 32.84 | 0.935 | 8.04 |
| Ours | **33.14** | **0.938** | **7.80** |

Results on UCF101

| | PSNR | SSIM | IE |
|---|---|---|---|
| w/o flow interpolation | 30.34 | 0.908 | 8.93 |
| w/o vis map | 31.16 | 0.918 | 8.33 |
| w/o perceptual loss | 30.96 | 0.916 | 8.50 |
| w/o warping loss | 30.52 | 0.910 | 8.80 |
| w/o smoothness loss | **31.19** | **0.918** | **8.26** |
| full model | **31.19** | **0.918** | 8.30 |

Ablation studies

# Real results



# Conclusions

1. Clever combination of old and new techniques can produce astonishing results.
2. When something goes wrong, just add another CNN.

# References

1. *Huaizu Jiang, Deqing Sun, Varun Jampani, Ming-Hsuan Yang, Erik G. Learned-Miller, Jan Kautz,* Super SloMo: High Quality Estimation of Multiple Intermediate Frames for Video Interpolation. http://arxiv.org/abs/1712.00080
2. *Olaf Ronneberger, Philipp Fischer, Thomas Brox,* U-Net: Convolutional Networks for Biomedical Image Segmentation. https://arxiv.org/abs/1505.04597