Multiframe Motion Coupling for Video Super Resolution

Jonas Geiping , Hendrik Dirks, Daniel Cremers, and Michael Moeller

Advances in Image Processing and Computer Vision Seminar

Chirag Bhuvaneshwara

2571703

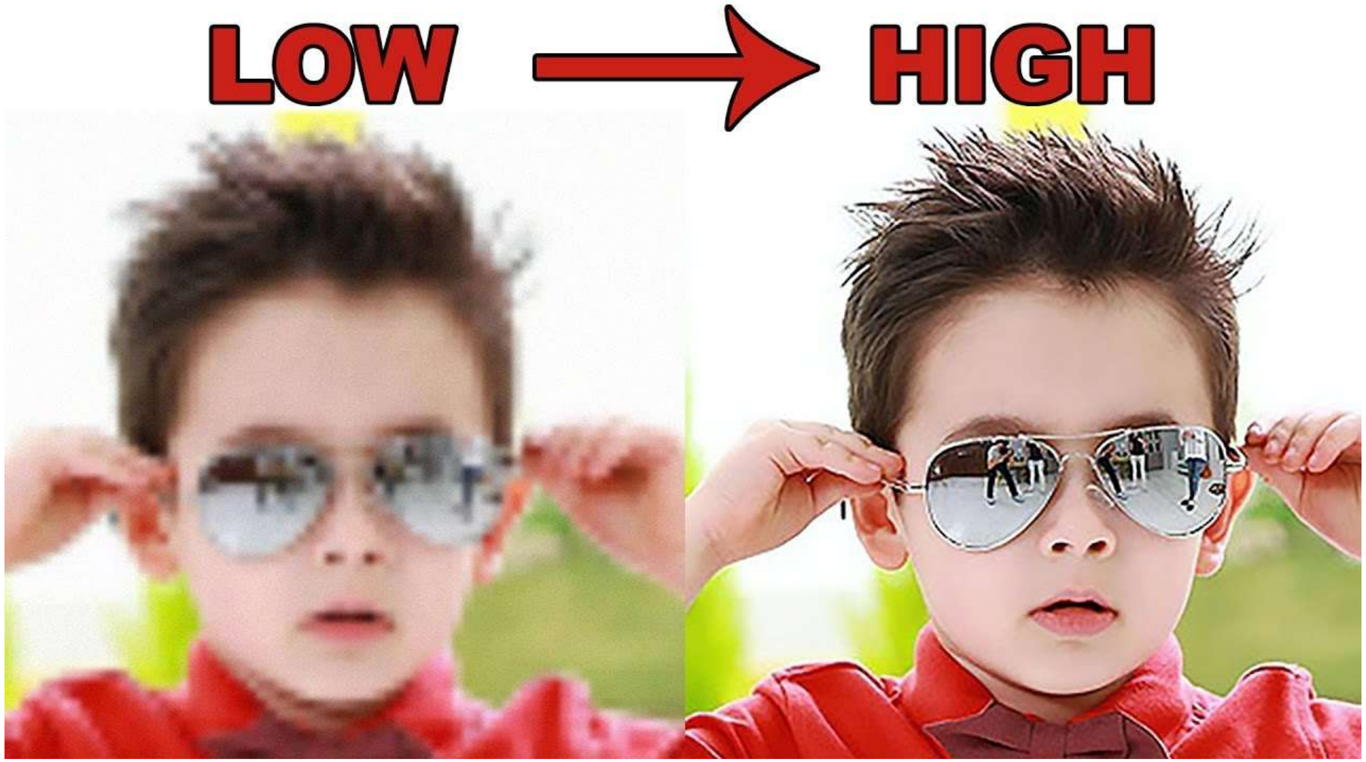
Universität des Saarlandes

Outline

1. Introduction
   - Goal
   - Techniques
2. Video Super Resolution
   - Classical Method
   - Proposed Method
3. MMC Model
   - The Full Model
   - Incremental Models
4. Experiments
   - Practical Considerations
   - Evaluation of Incremental Models
   - Comparison of Different Models
   - Sintel MPI Dataset Eval
5. Results
   - Example
   - Conclusion

Goal

- Need to upsample
- Some high freq info missing



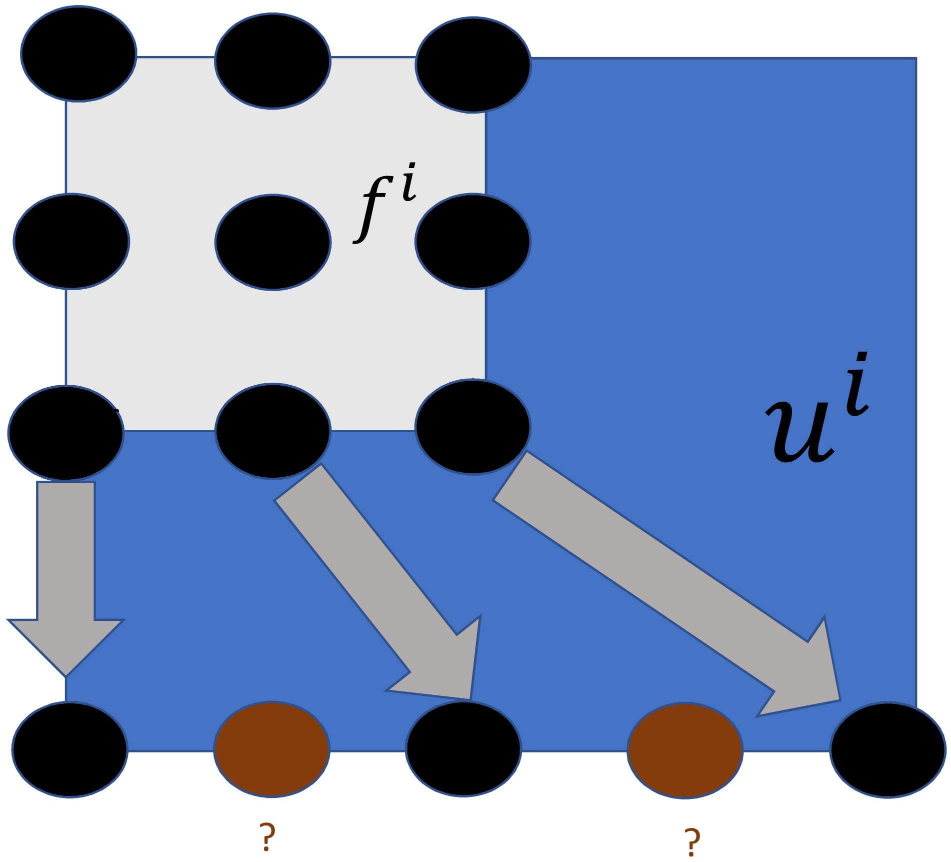
Source: youtube.com

Goal

Upsample notation

- f^i is low res frame (ex: 720p)

- u^i is high res frame (ex: 1080p)



Techniques

3 different techniques

1. Interpolation

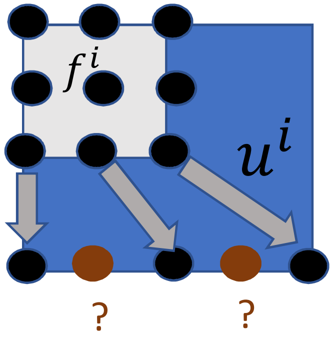
2. Variational Models

Proposed Model (MMC)

3. Deep Learning

DIFFERENT !

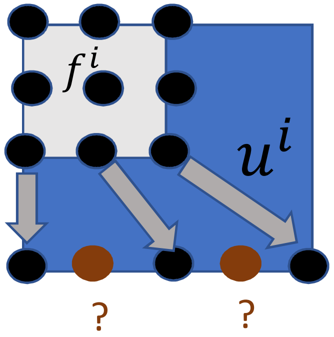
Perform Super Resolution



Techniques

Interpolation

- Nearest Neighbour Interpolation: Round co-ordinates of desired interpolation point. Find closest corresponding pixel in source image and use.

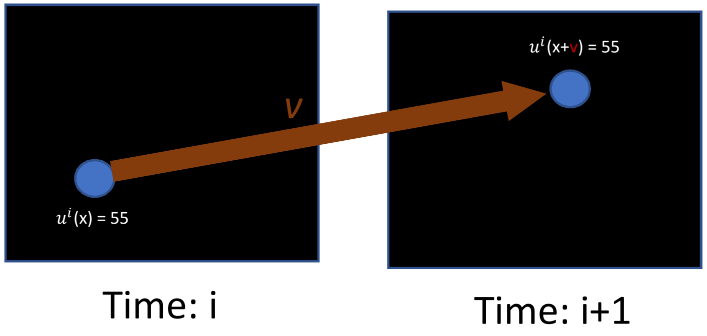


- Bicubic Interpolation: For each pixel, some sort of Gaussian weighting of the closest known pixels in the 4*4 grid surrounding it.

Techniques

Variational Video Super Resolution

- We have video.
- Compute Optic Flow v
- Develop a variational energy



$$u^* = \min_u E(u; f, v)$$

Techniques

Variational Video Super Resolution: The weighting

- Generally, in variational methods

$$E(u) = Data\ term + \alpha * Smoothness\ term$$

- In super resolution methods, in general, α is small.

$$\alpha < 1$$
$$\beta < 1$$
$$\kappa < 1$$

Techniques

Learning based Video Super Resolution

- CNNs are used.
- Both spatial and temporal information fed.
- Optic flow v is used in most high quality implementations.
- Ex: VSR Net, VDSR etc

Variational Video Super Resolution

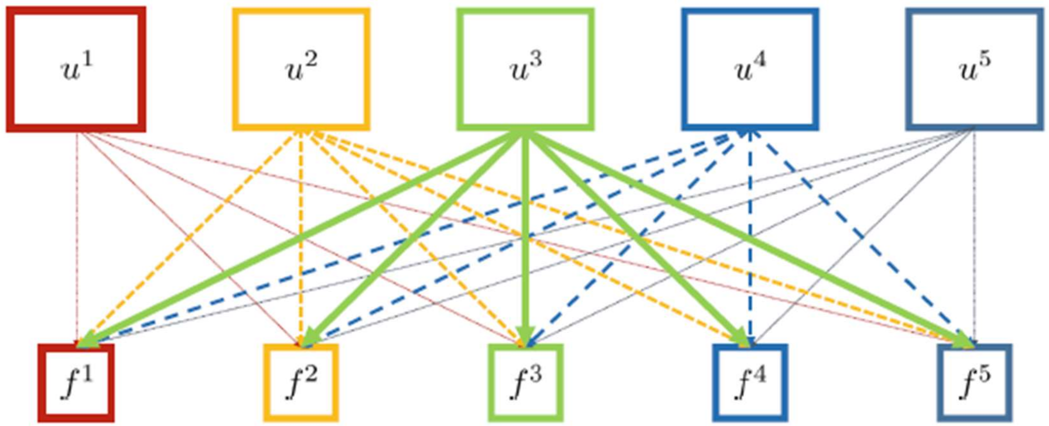
- Classical Methods
- Proposed Changes

Classical Methods

Coupling used

- To Compute 1 super resolved frame:

$$n\left\{\begin{array}{l} u^1 \propto \overbrace{\{OF(u^1, f^1), \ldots\ldots, OF(u^1, f^N)\}}^{n} \\ u^2 \propto \{OF(u^2, f^2), \ldots\ldots, OF(u^2, f^N)\} \\ \quad\quad\quad\quad \vdots \\ \quad\quad\quad\quad \vdots \end{array}\right.$$



Source: Geiping et al.

- No. of flow computations is $O(n^2)$ w.r.t no. of frames.

Classical Methods

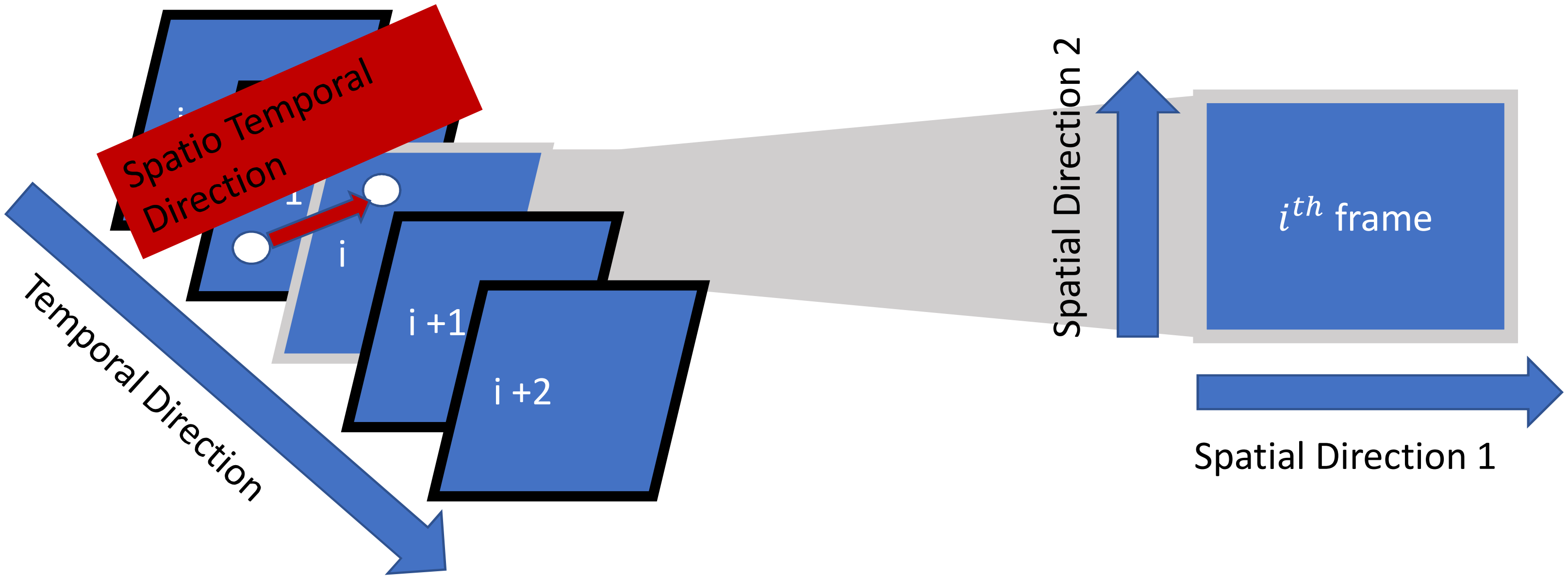
The Model: Some more notation

- *D is a downsampling operator* ➔ *dim*(Du^i) = *dim*(f^i)
- *b is a blurring operator* ➔ $b * u^i$ *removes high frequency info*
- *W is warping operator* ➔ *computes derivative in spatio temporal direction*

Classical Methods

The Model: The directions

- *W is warping operator ➔ computes derivative in spatio temporal direction*



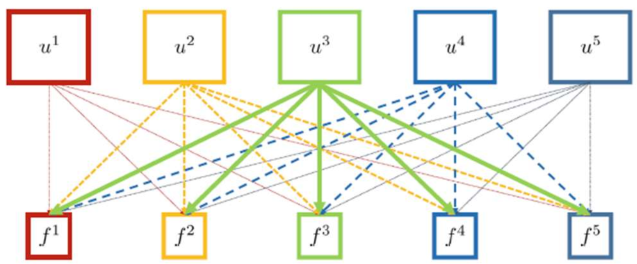
Classical Methods

The Model

➢Each super resolved frame is given as:

$$u^{i^*} = \min_{u^i} \underbrace{\|D(b * u^i) - f^i\|_{H^{\epsilon_d}}}_{\text{Spatial consistency data term}} + \lambda \underbrace{\|\nabla u^i\|_{H^{\epsilon_r}}}_{\text{Spatial consistency smoothness term}} + \sum_{j \neq i} \underbrace{\|D(b * W^{j,i} u^i) - f^j\|_{H^{\epsilon_d}}}_{\text{Temporal consistency data term}}$$

➢Temporal Consistency Smoothness term = 0



Source: Geiping et al.

Classical Methods

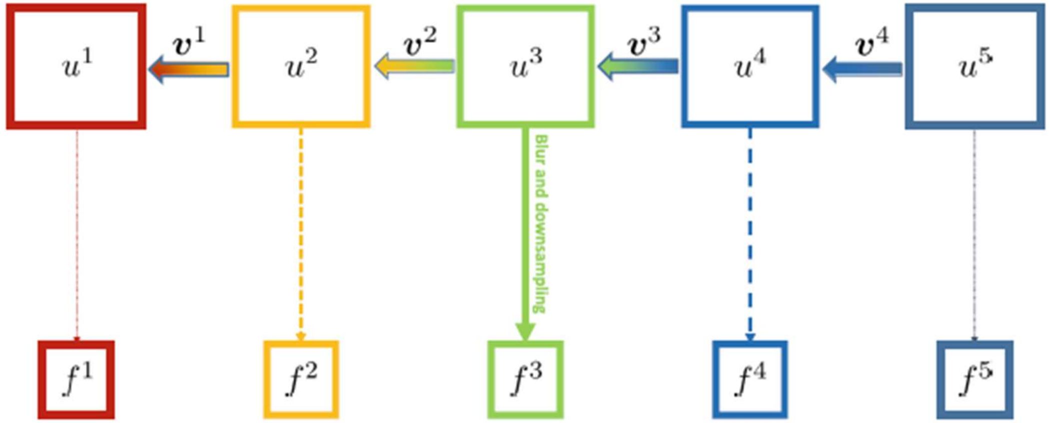
Disadvantages

- Too many OF computations
- May lead to flickering in super resolved video u

Proposed Changes
Multiframe Coupling

➢To Compute ALL super resolved Frames together:

$$n \left\{ \begin{array}{l} u^2 \propto \{OF(u^1, u^2)\} \\ u^3 \propto \{OF(u^2, u^3)\} \\ \vdots \\ \vdots \end{array} \right.$$



Source: Geiping et al.

➢No. of flow computations is $O(n)$ w.r.t no. of frames.

➢Brightness Constancy Assumption

Proposed Changes
Multiframe Coupling + Infimal Convolution

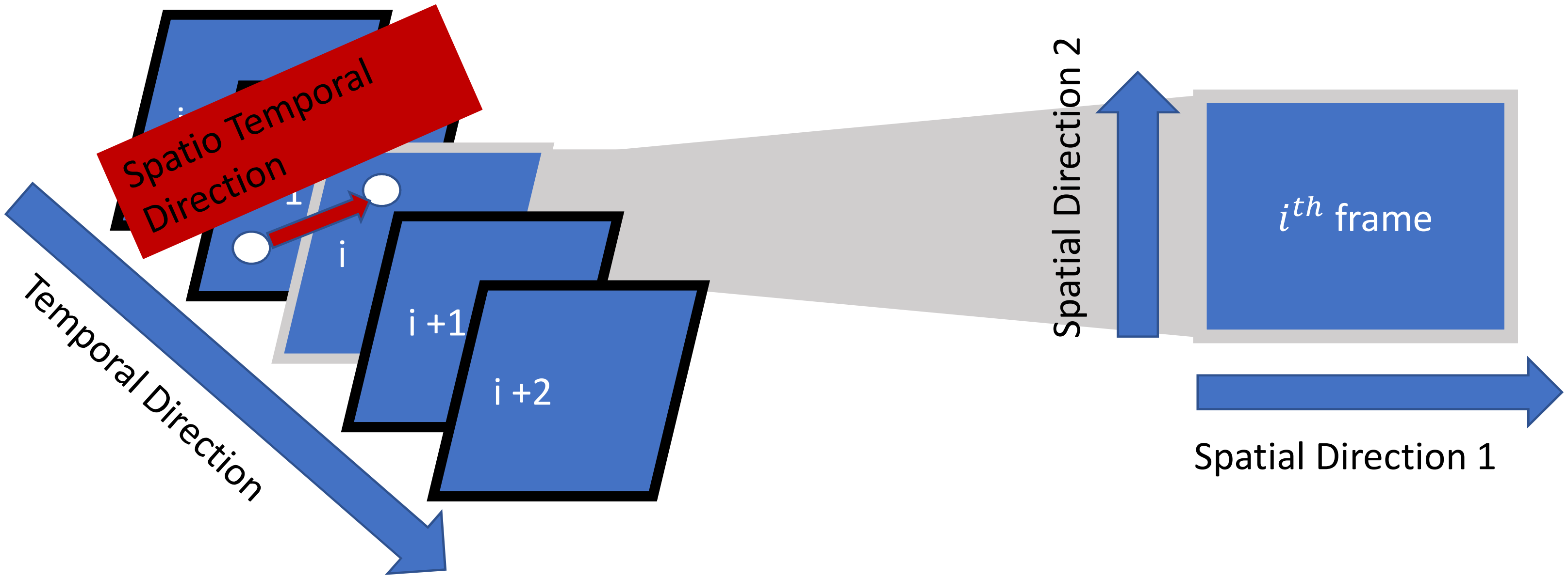
- $$u^* = \min_u \underbrace{\sum_{i=1}^{n} \|D(b * u^i) - f^i\|_1}_{\text{Spatial Consistency Data term}} + \alpha \underbrace{\inf_{u=w+z} R_{\text{temp}}(w) + R_{\text{spat}}(z).}_{\text{Infimal Convolution}}$$

- $$\alpha < 1$$

Proposed Changes

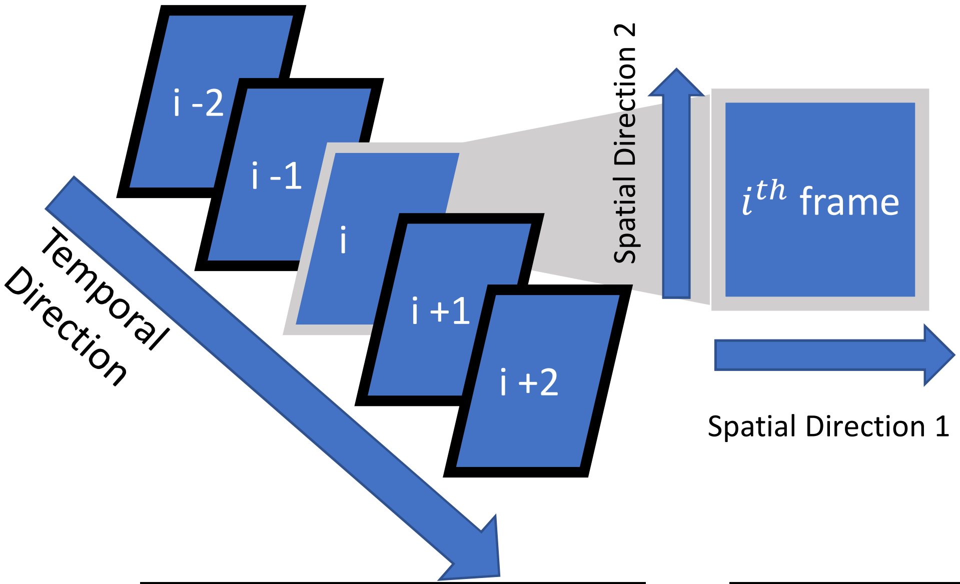
Infimal Convolution: The directions

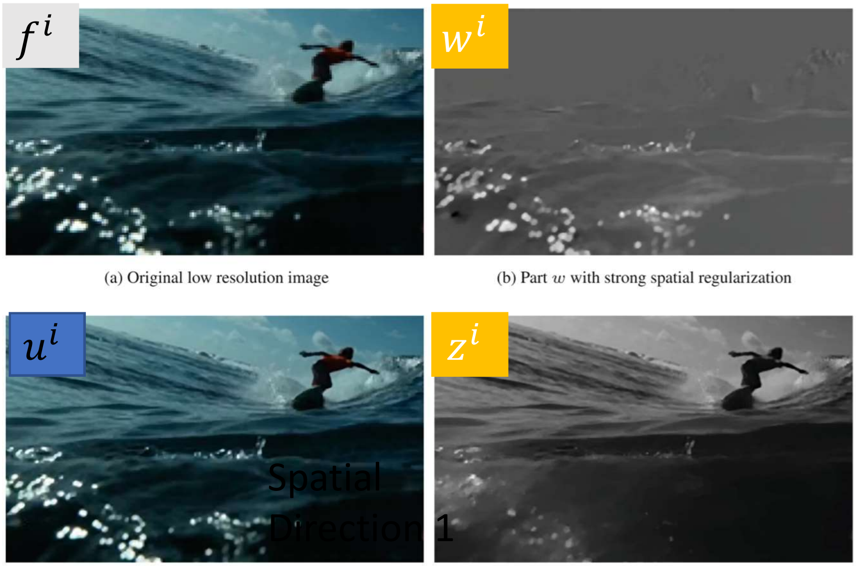
$$(R_{\mathrm{temp}} \Box R_{\mathrm{spat}})(u) := \inf_{u=w+z} R_{\mathrm{temp}}(w) + R_{\mathrm{spat}}(z).$$



Proposed Changes
Infimal Convolution

$$(R_{\text{temp}} \Box R_{\text{spat}})(u) := \inf_{u=w+z} \boxed{R_{\text{temp}}(w) + R_{\text{spat}}(z)}$$





(a) Original low resolution image

(b) Part w with strong spatial regularization

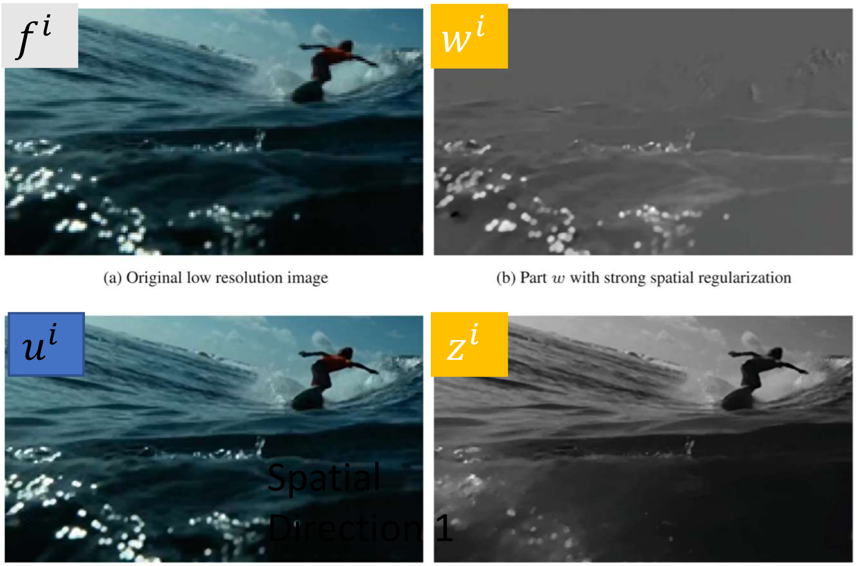
(c) Super resolution result u

(d) Part $u - w$ with strong temporal regularization

Source: Geiping et al.

Proposed Changes
Infimal Convolution

- $(R_{\text{temp}} \square R_{\text{spat}})(u) := \inf_{u=w+z} \boxed{R_{\text{temp}}(w) + R_{\text{spat}}(z)}$
- w^i ➔ OF accurate
- z^i ➔ OF not accuate
- $z = u - w$

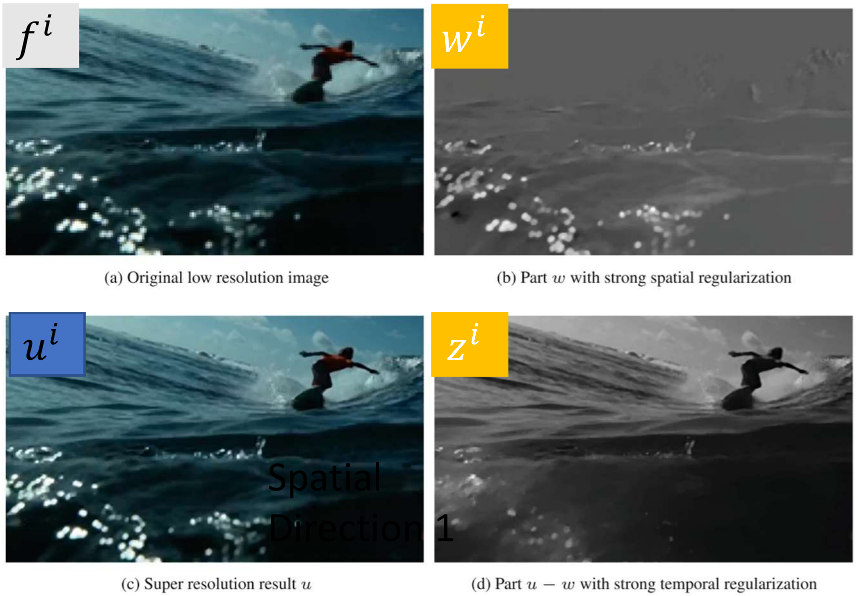


(a) Original low resolution image (b) Part w with strong spatial regularization (c) Super resolution result u (d) Part $u-w$ with strong temporal regularization

Source: Geiping et al.

Proposed Changes
Infimal Convolution

- $$(R_{\text{temp}} \square R_{\text{spat}})(u) := \inf_{u=w+z} \boxed{R_{\text{temp}}(w) + R_{\text{spat}}(z)}$$
- $R_{temp}(w) \rightarrow$ smooth w temporally (since we have OF)
- $R_{spat}(z) \rightarrow$ smooth z spatially
- $R_{spat}(z) = R_{spat}(u - w)$



(a) Original low resolution image

(b) Part w with strong spatial regularization

(c) Super resolution result u

(d) Part $u - w$ with strong temporal regularization

Source: Geiping et al.

Proposed Changes

Infimal Convolution: The convex functions

- $(R_{\text{temp}} \square R_{\text{spat}})(u) := \inf_{u=w+z} R_{\text{temp}}(w) + R_{\text{spat}}(z).$

- $R_{temp}(w) = \sum_{i=1}^{n} \left\| \sqrt{\underbrace{(\kappa w_x^i)^2 + (\kappa w_y^i)^2}_{\text{Spatial}} + \underbrace{W((w^i, w^{i+1}))^2}_{\text{Spatio temporal}}} \right\|_1$

- $R_{spat}(z) = \sum_{i=1}^{n} \left\| \sqrt{\underbrace{(z_x^i)^2 + (z_y^i)^2}_{\text{Spatial}} + \underbrace{\kappa W((z^i, z^{i+1}))^2}_{\text{Spatio temporal}}} \right\|_1$

$\kappa < 1$

Proposed Changes

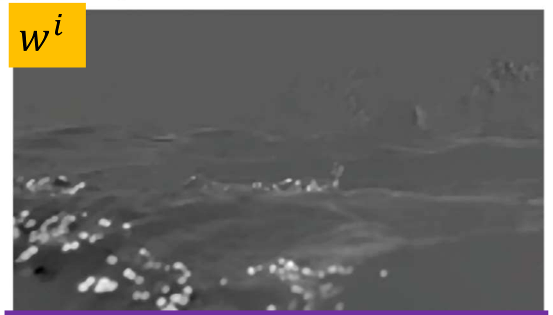
Infimal Convolution: The convex functions

- $(R_{\text{temp}} \square R_{\text{spat}})(u) := \inf_{u=w+z} R_{\text{temp}}(w) + R_{\text{spat}}(z).$

- $R_{temp}(w) \cong \sum_{i=1}^{n} \left\| \sqrt{\underbrace{W((w^i, w^{i+1}))^2}_{\text{Only Spatio temporal}}} \right\|_1$

- $R_{spat}(z) \cong \sum_{i=1}^{n} \left\| \sqrt{\underbrace{(z_x^i)^2 + (z_y^i)^2}_{\text{Only Spatial}}} \right\|_1$

Applying $\kappa < 1$



(b) Part w with strong spatial regularization

(d) Part $u - w$ with strong temporal regularization

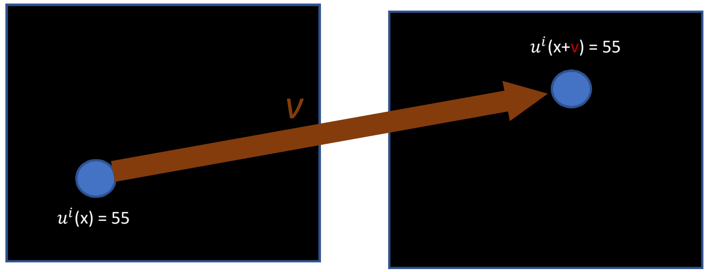
Source: Geiping et al.

Proposed Changes

Warping Operator: Motion Corrected time derivative

- $R_{temp}(w) = \sum_{i=1}^{n} \left\| \sqrt{(u_x^i)^2 + (u_y^i)^2 + \kappa \underbrace{W((u^i, u^{i+1}))^2}} \right\|_1$

- $$W((u^i, u^{i+1})) = \frac{\boldsymbol{u^i(x) - u^{i+1}(x + v^i(x))}}{\boldsymbol{h}}$$



Time: i

Time: i+1

Proposed Changes
Spatio Temporal Scaling

- $$W((u^i, u^{i+1})) = \frac{\boldsymbol{u^i(x) - u^{i+1}(x + v^i(x))}}{\boldsymbol{h}}$$

- $$\boldsymbol{h} = \frac{\|\boldsymbol{W u_0}\|_1}{\left\|\frac{\boldsymbol{d}}{\boldsymbol{dx}}\boldsymbol{u_0}\right\|_1 + \left\|\frac{\boldsymbol{d}}{\boldsymbol{dy}}\boldsymbol{u_0}\right\|_1}$$

$\boldsymbol{u_0}$ is bicubic interpolated video

- $\boldsymbol{W u_0}$ → vector valued image obtained by stacking $W((u^i, u^{i+1}))$, $\forall i$

MMC Model

- The Full Model
- Incremental Models

The Full Model

*Multiframe Motion Coupling (**MMC**)*

- $u^* = \min_{u} \sum_{i=1}^{n} \|D(b * u^i) - f^i\|_1 + \alpha \inf_{u=w+z} \boxed{R_{\text{temp}}(w) + R_{\text{spat}}(z)}$

- $R_{temp}(w) = \sum_{i=1}^{n} \left\| \sqrt{(\kappa w_x^i)^2 + (\kappa w_y^i)^2 + \underline{W((w^i, w^{i+1}))}^2} \right\|_1$

Warping Operator

- $W((u^i, u^{i+1})) = \frac{u^i(x) + u^{i+1}(x + v^i(x))}{\underline{h}}$

- $R_{spat}(z) = \sum_{i=1}^{n} \left\| \sqrt{(z_x^i)^2 + (z_y^i)^2 + \underline{\kappa W((z^i, z^{i+1}))}^2} \right\|_1$

Spatio Temporal Scaling

- $h = \frac{\|Wu_0\|_1}{\left\|\frac{d}{dx}u_0\right\|_1 + \left\|\frac{d}{dy}u_0\right\|_1}$

The Full Model

*Multiframe Motion Coupling (**MMC**): Compact notation*

- $$u^* = \arg\min_{u,w} \underbrace{\|\mathcal{A}u - f\|_1}_{\text{Spatial Consistency Data term}} + \alpha \underbrace{\left\| \begin{pmatrix} \nabla w \\ \kappa \mathbf{W} w \end{pmatrix} \right\|_{2,1}}_{R_{temp}(w)} + \alpha \underbrace{\left\| \begin{pmatrix} \kappa \nabla (u - w) \\ \mathbf{W}(u - w) \end{pmatrix} \right\|_{2,1}}_{R_{spat}(z) = R_{spat}(u-w)}$$

where $u = (u^1, \dots, u^n)$, $f = (f^1, \dots, f^n)$, and $\mathcal{A} = \mathrm{diag}(DB, \dots, DB)$

- $\boldsymbol{W w}$, $\boldsymbol{W}(\boldsymbol{u} - \boldsymbol{w})$ ➔ vector valued image obtained by stacking $W((u^i, u^{i+1}))$, $\forall i$

The Full Model

Optic Flow used

- $$\boldsymbol{v} = \arg\min_{\boldsymbol{v}} \sum_{i=1}^{n-1} \int_\Omega \overbrace{\|\nabla f^i(x) - \nabla f^{i+1}(x+\boldsymbol{v}^i(x))\|_1}^{\text{Data term 1: Gradient Constancy Assumption}} \, dx$$
$$+ \int_\Omega \underbrace{|f^i(x) - f^{i+1}(x+\boldsymbol{v}^i(x))|}_{\text{Data term 2: Brightness Constancy Assumption}} \, dx + \underset{\substack{\Big\downarrow \\ \beta < 1}}{\beta} \sum_{j=1}^{2} \underbrace{\|\nabla \boldsymbol{v}^i_j\|_{H^\epsilon}}_{\text{Smoothness Term}} .$$

The Full Model

Procedure

1. Compute OF on low res f .

2. Upsample OF to desired res by bicubic interpolation.

3. Solve Super Res problem & find u^*.

Incremental Models

Incremental Models
Additive Regularizer

Model 1

- $$u^* = \arg\min_u \underbrace{||\mathcal{A}u - f||_1}_{\substack{\text{Spatial Consistency} \\ \text{Data term}}} + \alpha \underbrace{||\mathcal{W}u||_1}_{\substack{\text{Temporal Consistency} \\ \text{Smoothness term}}} + \alpha \underbrace{||\nabla u||_{2,1}}_{\substack{\text{Spatial Consistency} \\ \text{Smoothness Term}}}$$

- *Spatio Temporal Scaling*

$$h = 1$$

Incremental Models

Additive Regularizer with Scaling

Model 2

- $$u^* = \arg\min_u \underbrace{||\mathcal{A}u - f||_1}_{\text{Spatial Consistency Data term}} + \alpha \underbrace{||\mathcal{W}u||_1}_{\text{Temporal Consistency Smoothness term}} + \alpha \underbrace{||\nabla u||_{2,1}}_{\text{Spatial Consistency Smoothness Term}}$$

- *Spatio Temporal Scaling*

$$\boldsymbol{h} = \frac{\|\boldsymbol{W u_0}\|_1}{\left\|\frac{\boldsymbol{d}}{\boldsymbol{dx}}\boldsymbol{u_0}\right\|_1 + \left\|\frac{\boldsymbol{d}}{\boldsymbol{dy}}\boldsymbol{u_0}\right\|_1}$$

Incremental Models
Infimal Convolution

Model 3

➢

$$u^* = \underset{u,w}{\arg\min} \underbrace{\|\mathcal{A}u - f\|_1}_{\text{Spatial Consistency Data term}} + \overbrace{\alpha \underbrace{\left\| \begin{pmatrix} \nabla w \\ \kappa \mathcal{W} w \end{pmatrix} \right\|_{2,1}}_{R_{temp}(w)} + \alpha \underbrace{\left\| \begin{pmatrix} \kappa \nabla (u - w) \\ \mathcal{W}(u - w) \end{pmatrix} \right\|_{2,1}}_{R_{spat}(z) = R_{spat}(u-w)}}^{\text{Infimal Convolution}}$$

➢ *Spatio Temporal Scaling*

$$h = 1$$

Incremental Models
Infimal Convolution with Scaling

Model 4

- Full MMC model

$$u^* = \underset{u,w}{\arg\min} \underbrace{\|\mathcal{A}u - f\|_1}_{\text{Spatial Consistency Data term}} + \alpha \overbrace{\underbrace{\left\| \begin{pmatrix} \nabla w \\ \kappa \mathcal{W} w \end{pmatrix} \right\|_{2,1}}_{R_{temp}(w)} + \alpha \underbrace{\left\| \begin{pmatrix} \kappa \nabla (u-w) \\ \mathcal{W}(u-w) \end{pmatrix} \right\|_{2,1}}_{R_{spat}(z) = R_{spat}(u-w)}}^{\text{Infimal Convolution}}$$

- *Spatio Temporal Scaling*

$$\boldsymbol{h} = \frac{\|\boldsymbol{W u_0}\|_1}{\left\|\frac{\boldsymbol{d}}{\boldsymbol{dx}}\boldsymbol{u_0}\right\|_1 + \left\|\frac{\boldsymbol{d}}{\boldsymbol{dy}}\boldsymbol{u_0}\right\|_1}$$

Experiments

- Practical Considerations
- Evaluation of Incremental Models
- Comparison of Different Models
- Sintel MPI Dataset Eval

Practical Considerations

- Convert to YCbCr, MMC on Y and Bicubic for Cb and Cr

- Longer Videos
  - Divide into frame batches.
  - Last frame from each batch as boundary condition.

- Best values for hyperparameters:

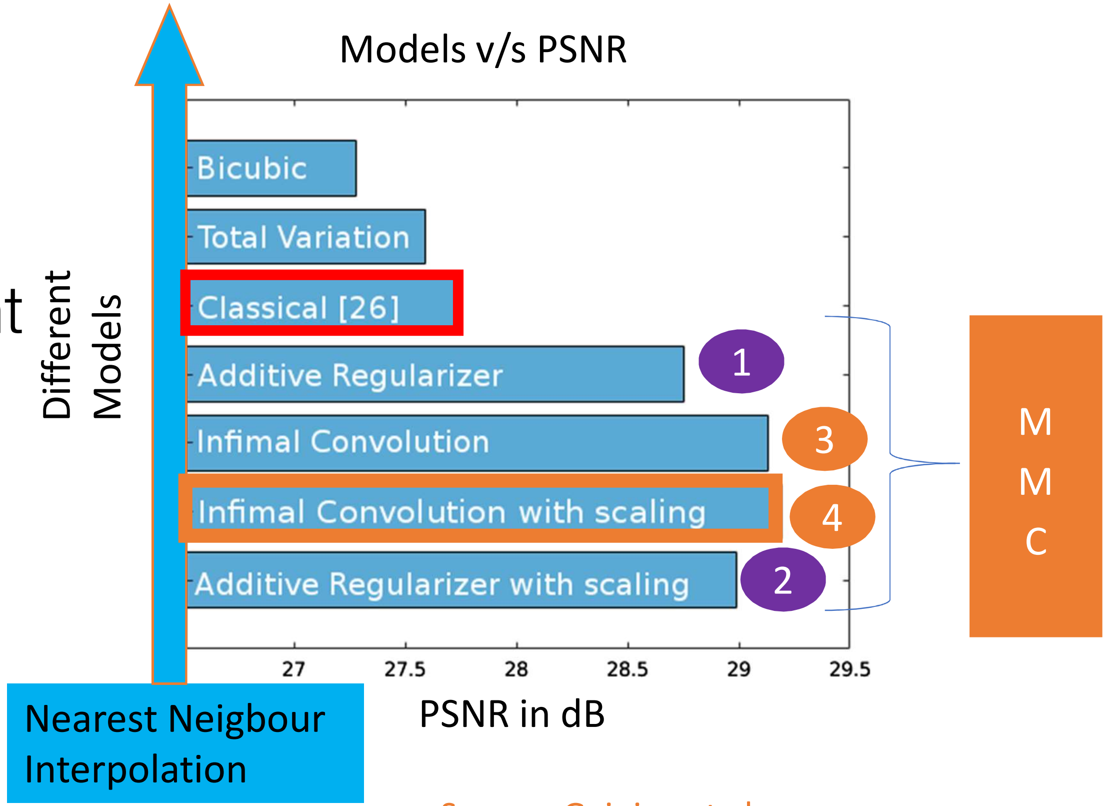
$$\boldsymbol{\beta = 0.2} \qquad \text{for Optic Flow}$$

$$\boldsymbol{\alpha = 0.01, \quad \kappa = 0.25} \qquad \text{for MMC Super Res}$$

Evaluation of Incremental Models

- Baseline: NN Interpolation

- Graph is showing improvement w.r.t baseline.

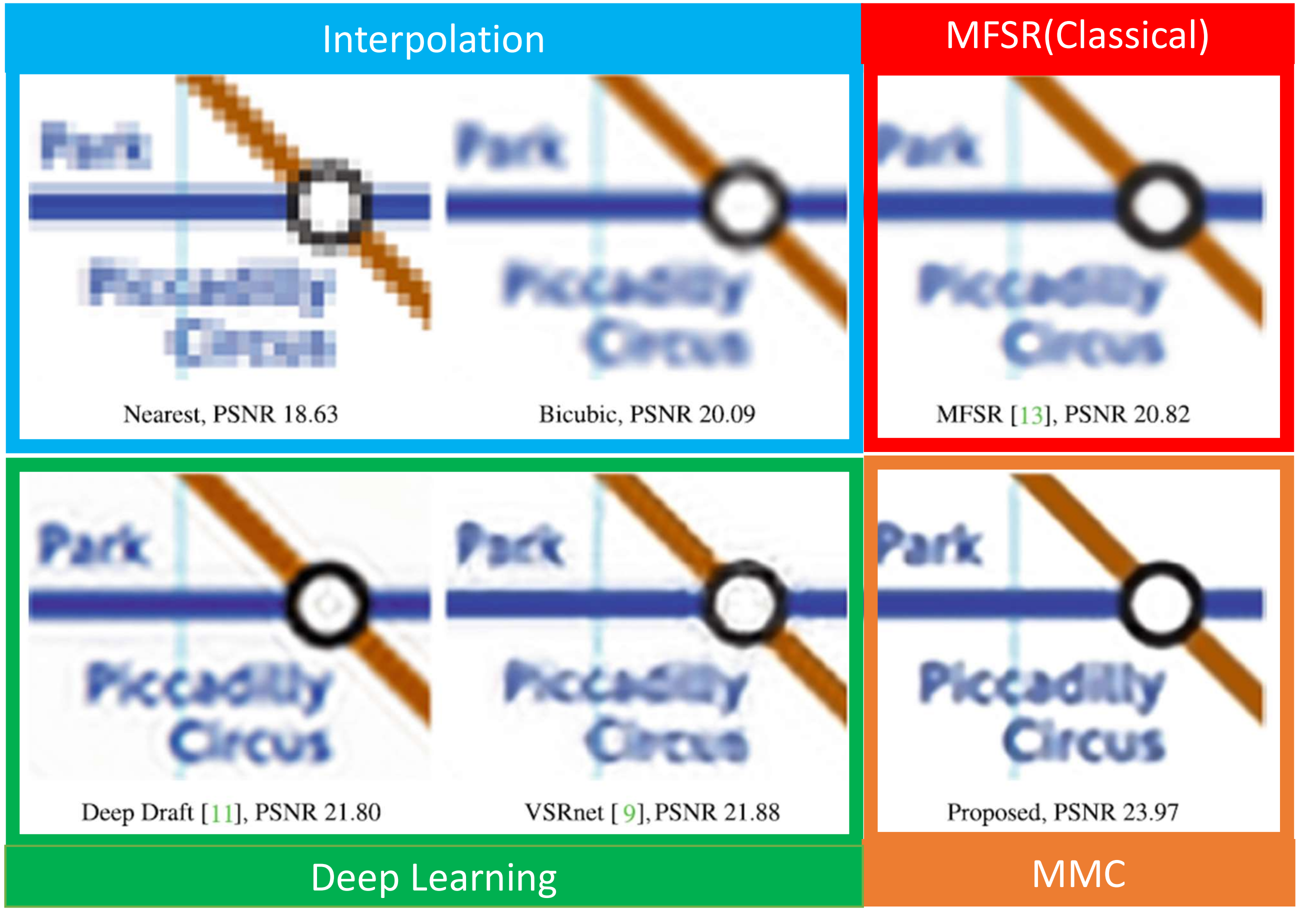


Source: Geiping et al.

Comparison of different Super Res Methods

Synthetic Dataset

- Ground Truth OF (GT OF) known
- Only planar motion



Source: Geiping et al.

Comparison of different Super Res Methods

Realistic Datasets

MMC (proposed method), PSNR 20.51 | MMC, PSNR 26.81 | MMC, PSNR 31.62 | MMC, 33.77

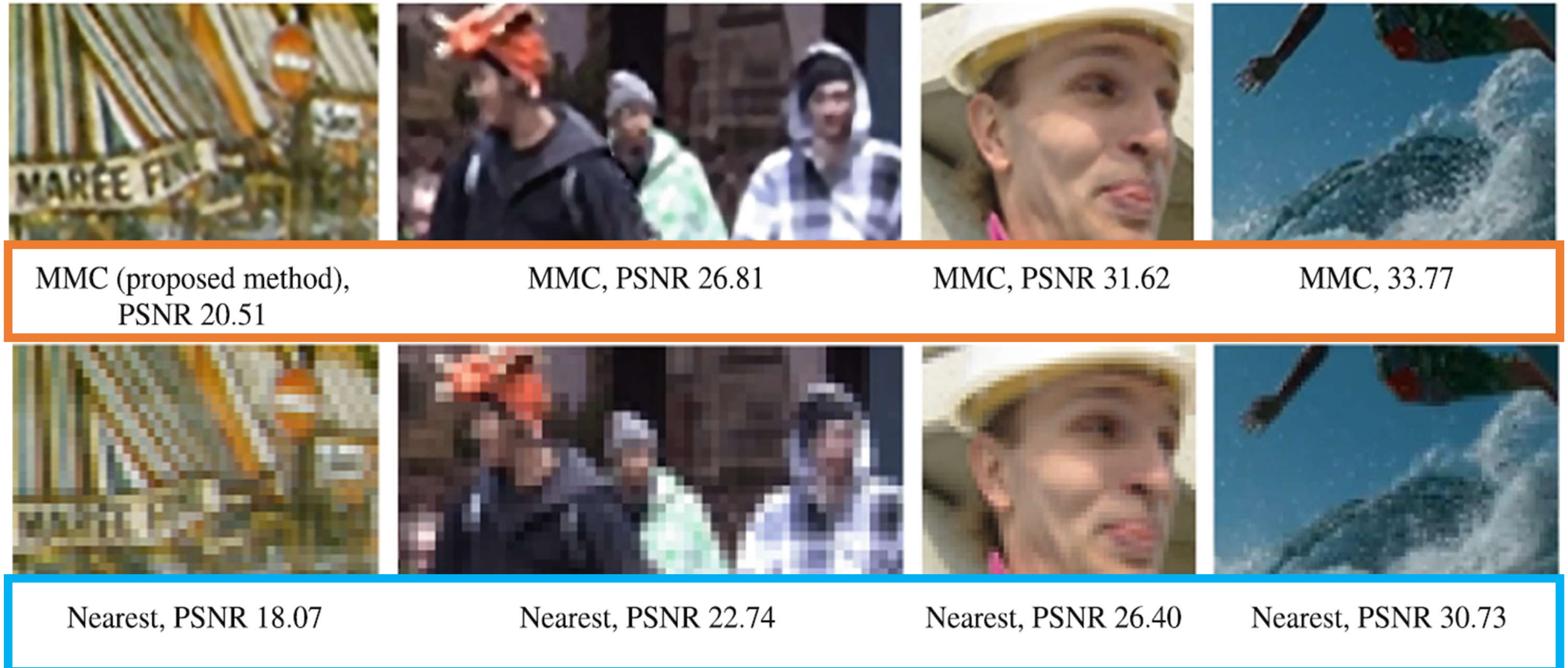
calendar dataset, ground truth zoom | *walk* dataset, gound truth zoom | *foreman* dataset, ground truth zoom | *wave* dataset, ground truth zoom

Ground Truth

Source: Geiping et al.

Comparison of different Super Res Methods

Realistic Datasets



Interpolation

Source: Geiping et al.

Comparison of different Super Res Methods

Realistic Datasets

MMC (proposed method), PSNR 20.51 | MMC, PSNR 26.81 | MMC, PSNR 31.62 | MMC, 33.77

MFSR [13], PSNR 19.20 | MFSR, PSNR 23.98 | MFSR, PSNR 28.39 | MFSR, PSNR 31.85

Variational Model (Classical Method)

Source: Geiping et al.

Comparison of different Super Res Methods

Realistic Datasets

MMC (proposed method), PSNR 20.51 | MMC, PSNR 26.81 | MMC, PSNR 31.62 | MMC, 33.77

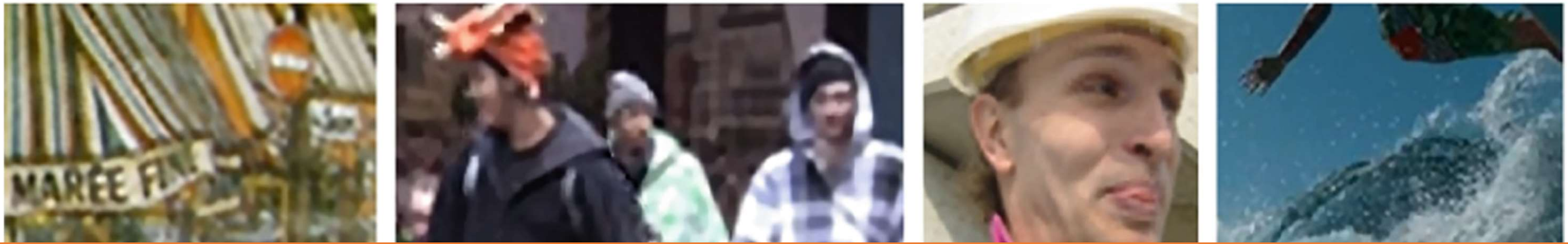
VSRnet [9], PSNR 19.36 | VSRnet, PSNR 25.95 | VSRnet, PSNR 31.02 | VSRnet, PSNR 33.03

Deep Learning

Source: Geiping et al.

Comparison of different Super Res Methods

Realistic Datasets



MMC (proposed method), PSNR 20.51 | MMC, PSNR 26.81 | MMC, PSNR 31.62 | MMC, 33.77

VDSR [10], PSNR 19.63 | VDSR, PSNR 26.40 | VDSR, PSNR 32.54 | VDSR, PSNR 33.33

Deep Learning

Source: Geiping et al.

Comparison of different Super Res Methods

Realistic Datasets

	Interpolation			Variational			MMC	Deep Learn		
PSNR	NN	BIC	VidEn	Mitxel	Unger	MFSR	MMC	DD	VSRNet	VDSR
Tube	18.63	20.09	21.73	23.57	23.11	20.82	**23.97**	21.80	21.88	22.36
City	23.35	23.95	24.75	25.38	25.14	24.23	**25.57**	24.92	24.45	24.60
Calendar	18.07	18.71	19.49	20.45	20.13	19.20	**20.51**	19.91	19.36	19.63
Foliage	21.21	22.21	23.19	23.41	23.38	22.40	**24.25**	23.45	23.00	23.16
Walk	22.74	24.37	25.37	24.29	24.13	23.98	**26.81**	25.00	25.95	26.40
Foreman	26.40	28.66	29.31	29.51	29.13	28.39	31.62	28.95	31.02	**32.54**
Temple	24.15	25.47	26.29	26.79	26.76	25.84	27.66	25.35	27.39	**27.90**
Penguins	29.17	31.77	32.82	32.55	32.78	32.54	32.91	30.56	34.63	**35.00**
Sheets	29.68	32.76	33.73	33.49	33.55	32.27	**34.23**	33.01	33.86	33.85
Surfer	30.59	32.91	33.29	26.52	27.15	29.11	34.42	30.48	30.45	**34.96**
Wave	30.73	31.96	32.82	32.81	32.81	31.85	**33.77**	32.43	33.03	33.33
Dog	32.58	34.48	35.07	34.54	34.77	34.15	35.18	34.09	35.63	**35.71**
Average	25.61	27.28	28.16	27.77	27.74	27.07	**29.19**	27.50	28.39	29.13

Source: Geiping et al.

Sintel MPI Dataset Eval

Synthetic Dataset

- ➢Increasing levels of Realism
- ➢Contains GT OF
- ➢Our OF Assumptions & MMC Assumptions are in sync

on Sintel [4] dataset **bandage_1**

More Realsitic

Rendering	GT Flow	Our OF
Albedo	32.53	31.91
Clean	27.88	27.68
Final	33.31	34.65

Source: Geiping et al.

Results

- Example Video
- Conclusion

Example Video

Ground Truth | Interpolation

Deep Learning | MMC

Source: Geiping et al. associated website

Conclusion

- MMC better than classical methods as it removes flickering.
- Reasonable motion/ high frame rate => MMC better than DL.
- Large motion/strong occlusions => DL better than MMC.
  But MMC is still good enough.

References

- Geiping, Jonas & Dirks, Hendrik & Cremers, Daniel & Moeller, Michael. (2018). Multiframe Motion Coupling for Video Super Resolution. 10.1007/978-3-319-78199-0_9., associated website: http://www.vsa.informatik.uni-siegen.de/en/superResolution
- Parsania, Pankaj & Paresh V. Virparia, Dr. (2016). A Comparative Analysis of Image Interpolation Algorithms. IJARCCE. 5. 29-34. 10.17148/IJARCCE.2016.5107.
- Holler, M., Kunisch, K.: On infimal convolution of TV-type functionals and applications to video and image reconstruction. SIAM J. Imaging Sci. 7(4), 2258–2300 (2014)
- Unger, M., Pock, T., Werlberger, M., Bischof, H.: A convex approach for variational super-resolution. In: Goesele, M., Roth, S., Kuijper, A., Schiele, B., Schindler, K. (eds.) DAGM 2010. LNCS, vol. 6376, pp. 313–322. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-15986-2 32
- J. Weickert, Image Processing & Computer Vision lecture 2018