JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

Paper

HyperCUT: Video Sequence From a Single Blurry Image Using Unsupervised Ordering

Bang-Dang Pham[1,*]

Phong Tran[1,2,*]

Anh Tran[1]

Cuong Pham[1,3]

Rang Nguyen[1]

Minh Hoai[1,4]

[1]VinAI Research, Vietnam [2]MBZUAI, UAE [3]Posts & Telecommunications Inst. of Tech., Vietnam [4]Stony Brook University, USA

[*]*Equal contribution*

Poster: **WED-AM-155**

Quick Preview

Image-to-video deblurring (**Blur2Vid**)

Blur input

Seq. of sharp images

Quick Preview

JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

Image-to-video deblurring (**Blur2Vid**)

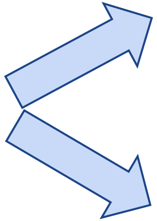
Blur input

Seq. of sharp images

Quick Preview

Image-to-video deblurring (**Blur2Vid**)

Blur input



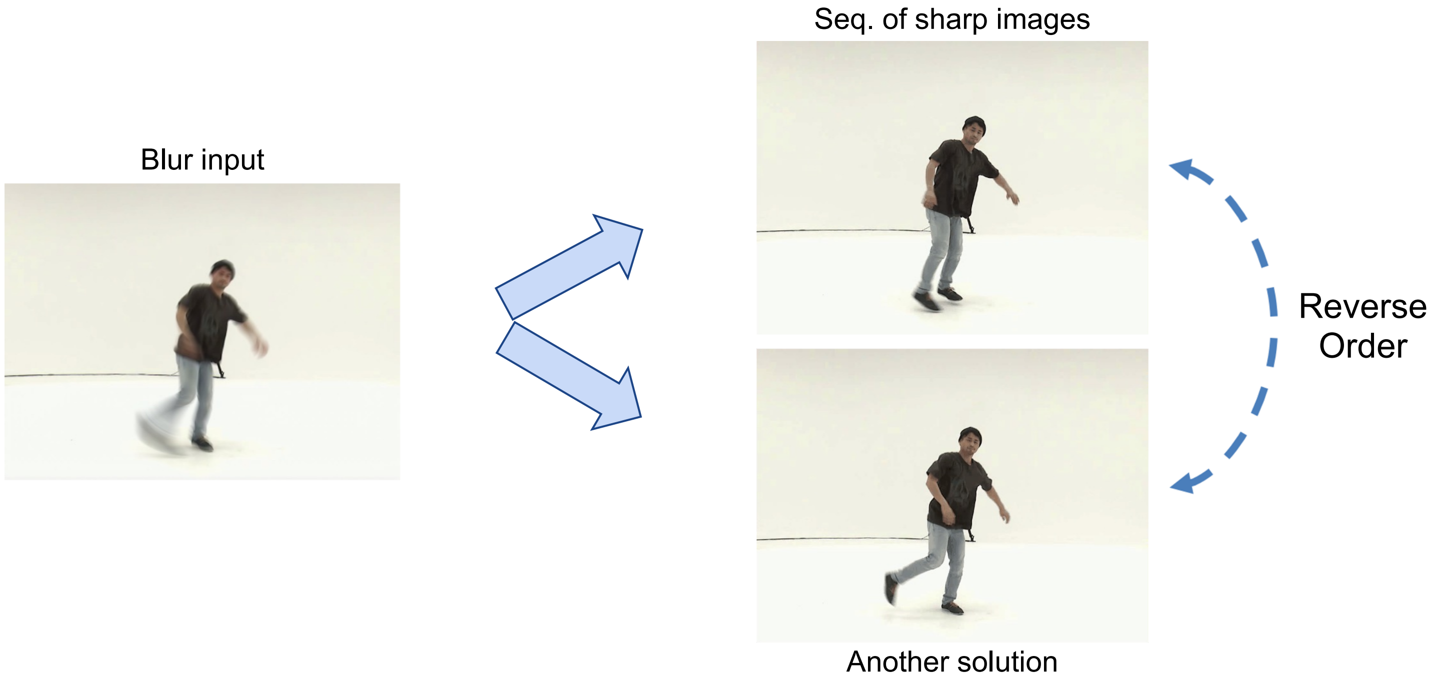
Seq. of sharp images

Another solution

Quick Preview

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

Image-to-video deblurring (**Blur2Vid**)



Quick Preview

Order-Ambiguity Issue

Blur input

Seq. of sharp images

Reverse Order

Another solution

Quick Preview

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

Order-Ambiguity Issue

Blur input

Seq. of sharp images

Another solution

Reverse Order

Quick Preview

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

Order-Ambiguity Issue ⇨ Fail Training

Blur input

Quick Preview

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

Order-Ambiguity Issue ⇨ Fail Training ⇨ Incorrect Motion

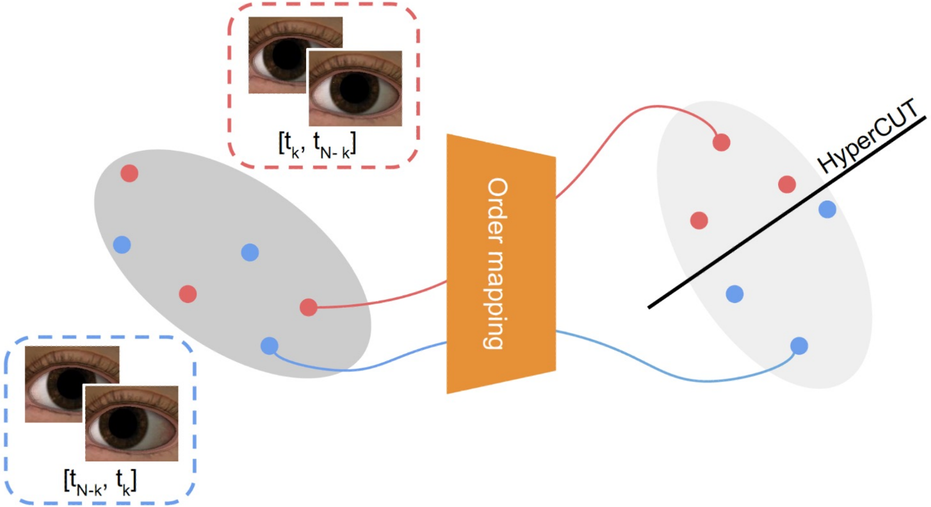
Blur input

Failure Generated Frames

Quick Preview

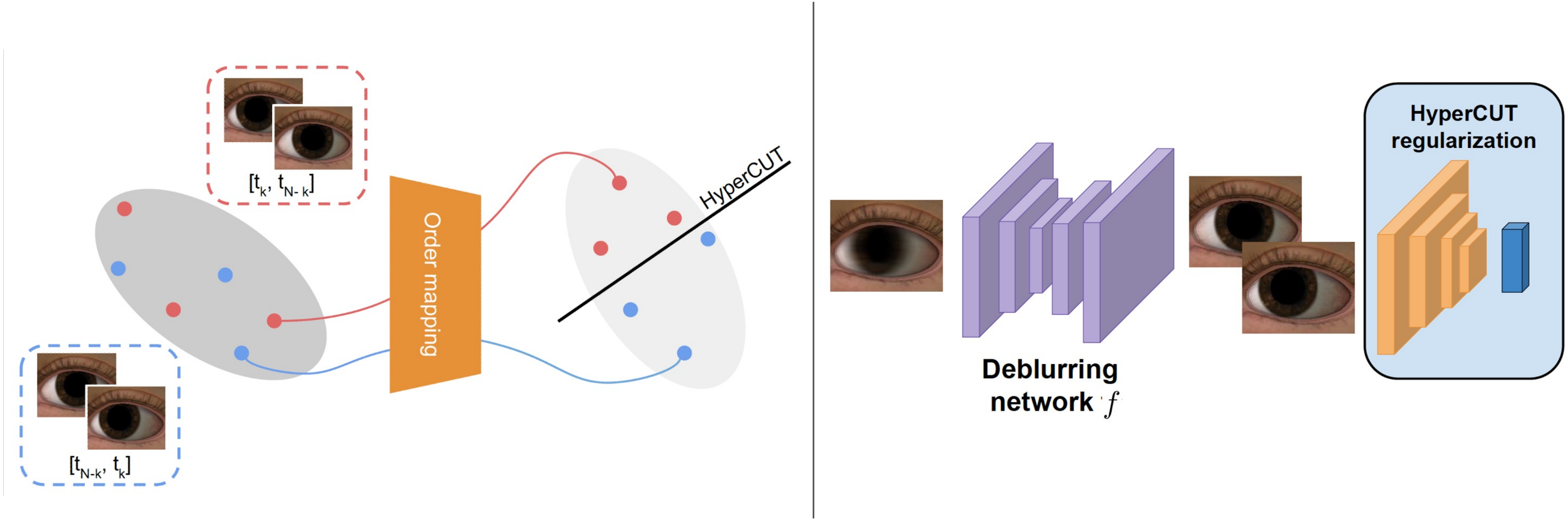
JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

- We propose **HyperCUT**, an effective **Self-supervised Ordering Scheme** that **assigns** an **explicit order** for each video sequence, thus avoiding the **order-ambiguity issue**



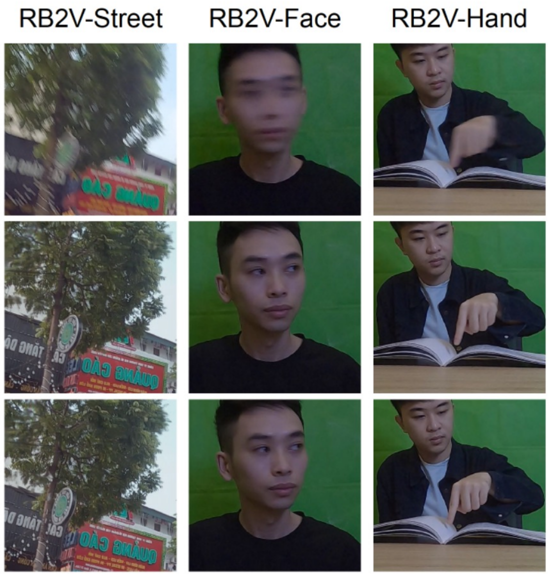
Quick Preview

- We propose **HyperCUT**, an effective **Self-supervised Ordering Scheme** that **assigns** an **explicit order** for each video sequence, thus avoiding the **order-ambiguity issue**



Quick Preview

- We propose **HyperCUT**, an effective **Self-supervised Ordering Scheme** that **assigns** an **explicit order** for each video sequence, thus avoiding the **order-ambiguity issue**
- We provide the **first** **real-world blur2vid** dataset covers a variety of popular domains, including **face**, **hand**, and **street**



Data subset	#data samples	
	Train	Test
RB2V-Street	9000	2053
RB2V-Face	8000	2157
RB2V-Hand	12000	4722

Dataset Statistics

Quick Preview

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

- We propose **HyperCUT**, an effective **Self-supervised Ordering Scheme** that **assigns** an **explicit order** for each video sequence, thus avoiding the **order-ambiguity issue**
- We provide the **first real-world blur2vid** dataset covers a variety of popular domains, including **face**, **hand**, and **street**

Generated frames

Baselines

Model		1^{st}	2^{nd}	3^{rd}	4^{th}	5^{th}	6^{th}	7^{th}
[1]	REDS	20.65	22.63	24.20	23.50	24.20	22.63	20.65
[1] + HyperCUT	REDS	**22.87**	**24.88**	**26.29**	**25.10**	**26.29**	**24.88**	**22.86**
[2]	REDS	22.78	24.47	26.14	**31.50**	26.12	24.49	22.83
[2] + HyperCUT	REDS	**26.75**	**28.30**	**29.42**	29.97	**29.41**	**28.30**	**26.76**
[2]	RB2V	26.99	27.99	29.45	**32.08**	29.55	28.06	27.04
[2] + HyperCUT	RB2V	**28.29**	**29.20**	**30.43**	**32.08**	**30.53**	**29.22**	**28.25**

[1] Meiguang Jin, Givi Meishvili, and Paolo Favaro. Learning to extract a video sequence from a single motion-blurred image. CVPR, 2018
[2] Kuldeep Purohit, Anshul Shah, and AN Rajagopalan. Bringing alive blurred moments. CVPR, 2019

Quick Preview

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

- We propose **HyperCUT**, an effective **Self-supervised Ordering Scheme** that **assigns** an **explicit order** for each video sequence, thus avoiding the **order-ambiguity issue**
- We provide the **first real-world blur2vid** dataset covers a variety of popular domains, including **face**, **hand**, and **street**

Generated frames

Baselines

Model		1^{st}	2^{nd}	3^{rd}	4^{th}	5^{th}	6^{th}	7^{th}
[1]	REDS	20.65	22.63	24.20	23.50	24.20	22.63	20.65
[1] + HyperCUT	REDS	**22.87**	**24.88**	**26.29**	**25.10**	**26.29**	**24.88**	**22.86**
[2]	REDS	22.78	24.47	26.14	**31.50**	26.12	24.49	22.83
[2] + HyperCUT	REDS	**26.75**	**28.30**	**29.42**	29.97	**29.41**	**28.30**	**26.76**
[2]	RB2V	26.99	27.99	29.45	**32.08**	29.55	28.06	27.04
[2] + HyperCUT	RB2V	**28.29**	**29.20**	**30.43**	**32.08**	**30.53**	**29.22**	**28.25**

[1] Meiguang Jin, Givi Meishvili, and Paolo Favaro. Learning to extract a video sequence from a single motion-blurred image. CVPR, 2018
[2] Kuldeep Purohit, Anshul Shah, and AN Rajagopalan. Bringing alive blurred moments. CVPR, 2019

Motivation

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

Blur

Deblur

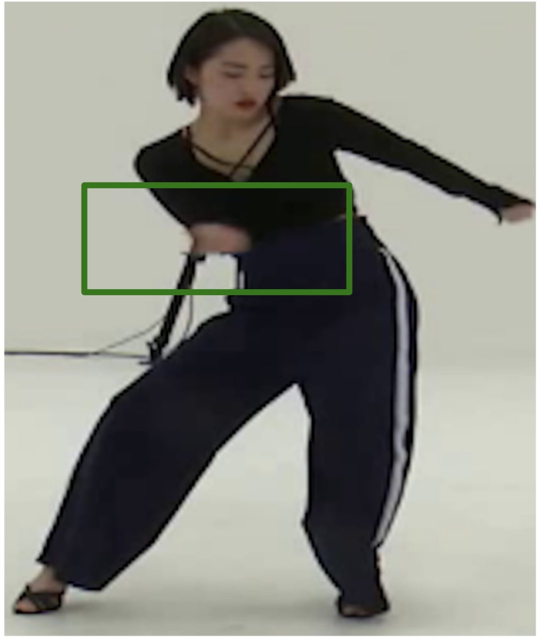
Sharp

Motivation

Blur

Deblur

Sharp



Motivation

Blur

Deblur

Sharp images

Motivation

Motivation

Deblur

Better Motion Analysis

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

Previous Approaches

Previous Approaches

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

- Naive Optimization approach:

$$f_k = \underset{f}{\operatorname{argmin}}\, \mathbb{E}_{x,y}||f(y) - x_k||_2^2$$

with y: blur input

x_k: ground-truth sharp images

f_k: network predict each sharp frame

Previous Approaches

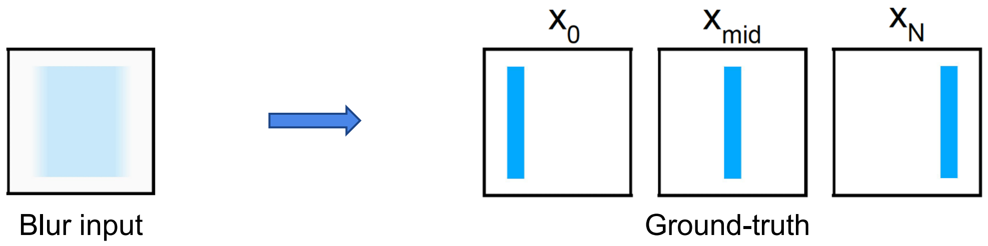
- Naive Optimization approach:

$$f_k = \underset{f}{\operatorname{argmin}} \, \mathbb{E}_{x,y} ||f(y) - x_k||_2^2$$

with y : blur input

x_k: ground-truth sharp images

f_k: network predict each sharp frame



Previous Approaches

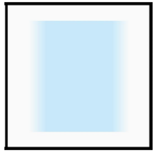
- Naive Optimization approach:

$$f_k = \underset{f}{\operatorname{argmin}} \, \mathbb{E}_{x,y} ||f(y) - x_k||_2^2$$

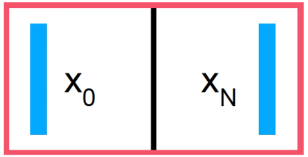
with y: blur input

x_k: ground-truth sharp images

f_k: network predict each sharp frame



Blur input



Ground-truth

Previous Approaches

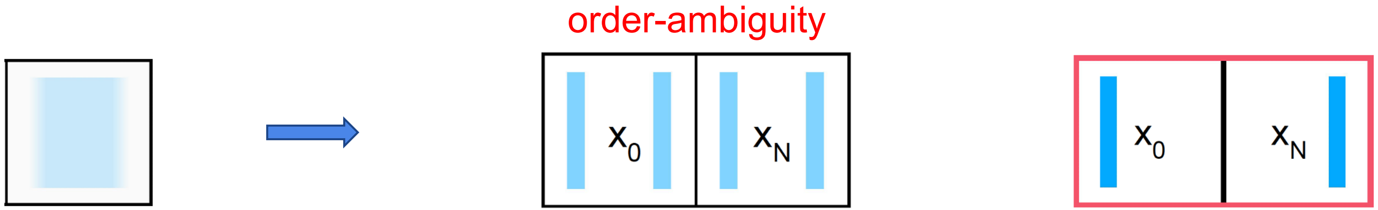
- Naive Optimization approach:

$$f_k = \underset{f}{\operatorname{argmin}}\, \mathbb{E}_{x,y} ||f(y) - x_k||_2^2$$

with y : blur input

x_k: ground-truth sharp images

f_k: network predict each sharp frame



Blur input Naive deblurring Ground-truth

Previous Approaches

- Order-Invariant loss:

$$\mathcal{L}_{OI} = \sum_{k=0}^{2} \left(\big| \|f_k(y) - f_{6-k}(y)\| - \|x_k - x_{6-k}\| \big| + \big| \|f_k(y) + f_{6-k}(y)\| - \|x_k + x_{6-k}\| \big| \right)$$

$\longrightarrow$ accept any order

with y : blur input

x_k: ground-truth sharp images $(x_0, \ldots, x_6)$

f_k: network predict each sharp frame $(f_0, \ldots, f_6)$

Previous Approaches

- Order-Invariant loss:

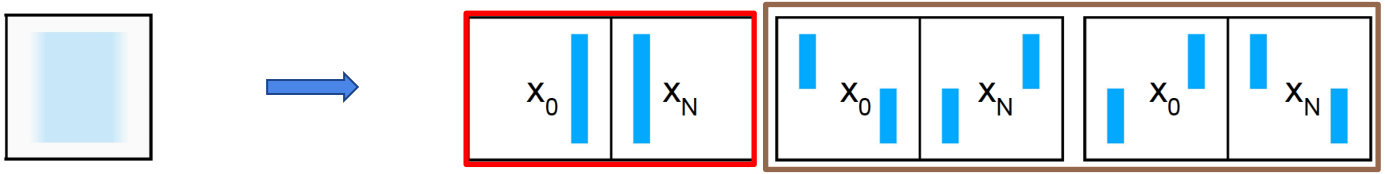
$$\mathcal{L}_{OI} = \sum_{k=0}^{2} \left(\big| \|f_k(y) - f_{6-k}(y)\| - \|x_k - x_{6-k}\| \big| + \big| \|f_k(y) + f_{6-k}(y)\| - \|x_k + x_{6-k}\| \big| \right)$$

→ accept any order

with y: blur input

x_k: ground-truth sharp images $(x_0, \ldots, x_6)$

f_k: network predict each sharp frame $(f_0, \ldots, f_6)$

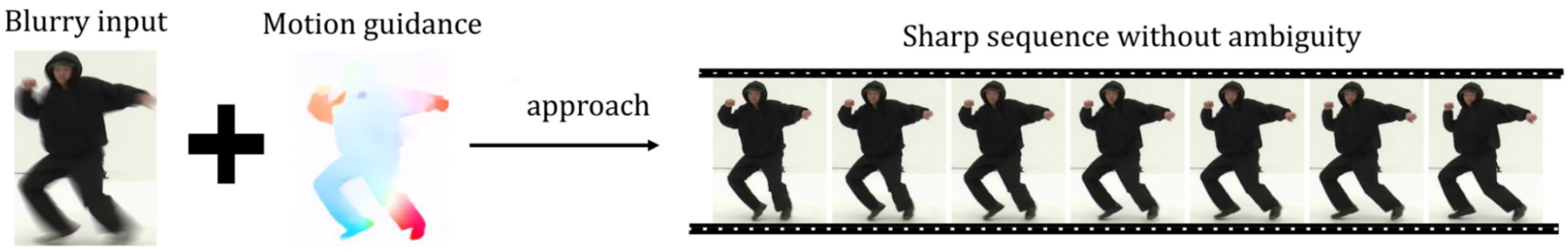


Blur input

Possible Solutions

Previous Approaches

- Motion Guidance:



✅ Multi-directional motion

✅ Refinement

❌ Accurate annotation requirement

❌ Handcrafted heuristics dependence

Figure from Zhong et al. “Animation from blur: Multimodal blur decomposition with motion guidance”. ECCV, 2022

Previous Approaches

- Motion Guidance:

Our goal

Separate the solution space

Consistent and efficient order alignment

No heuristic dependence

✅ Multi-directional motion

✅ Refinement

❌ Accurate annotation requirement

❌ Handcrafted heuristics dependence

Proposed Method

$[x_k, x_{N-k}]$

$[x_{N-k}, x_k]$

Proposed Method

$[x_k, x_{N-k}]$

$[x_{N-k}, x_k]$

Proposed Method

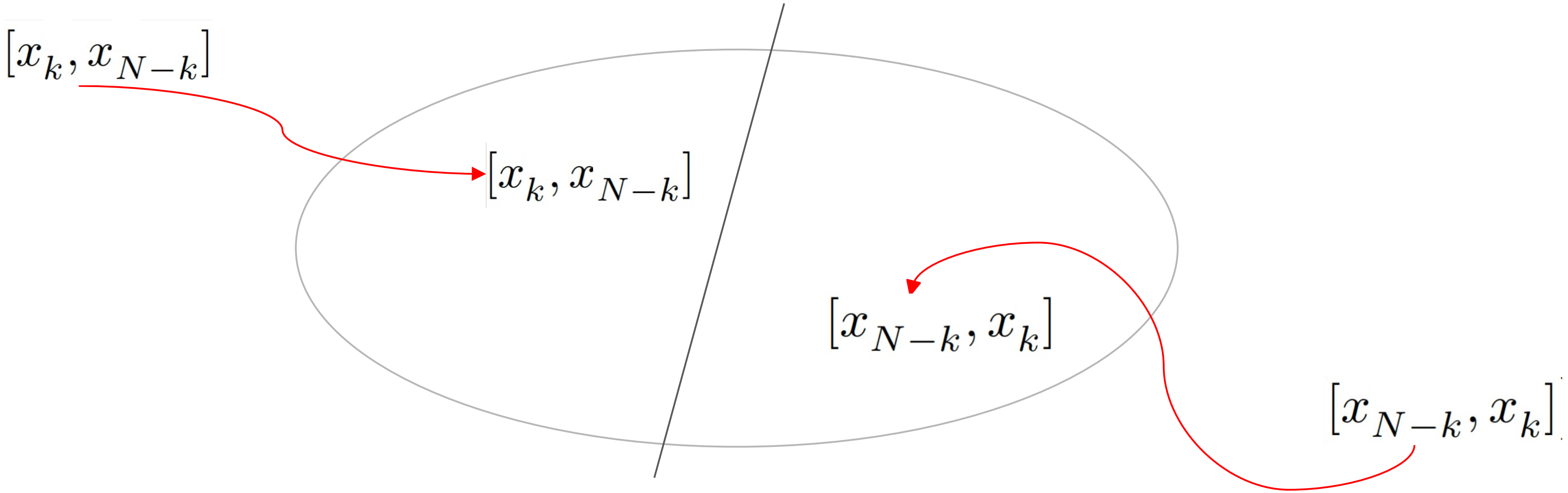
$[x_k, x_{N-k}]$

$[x_k, x_{N-k}]$

$[x_{N-k}, x_k]$

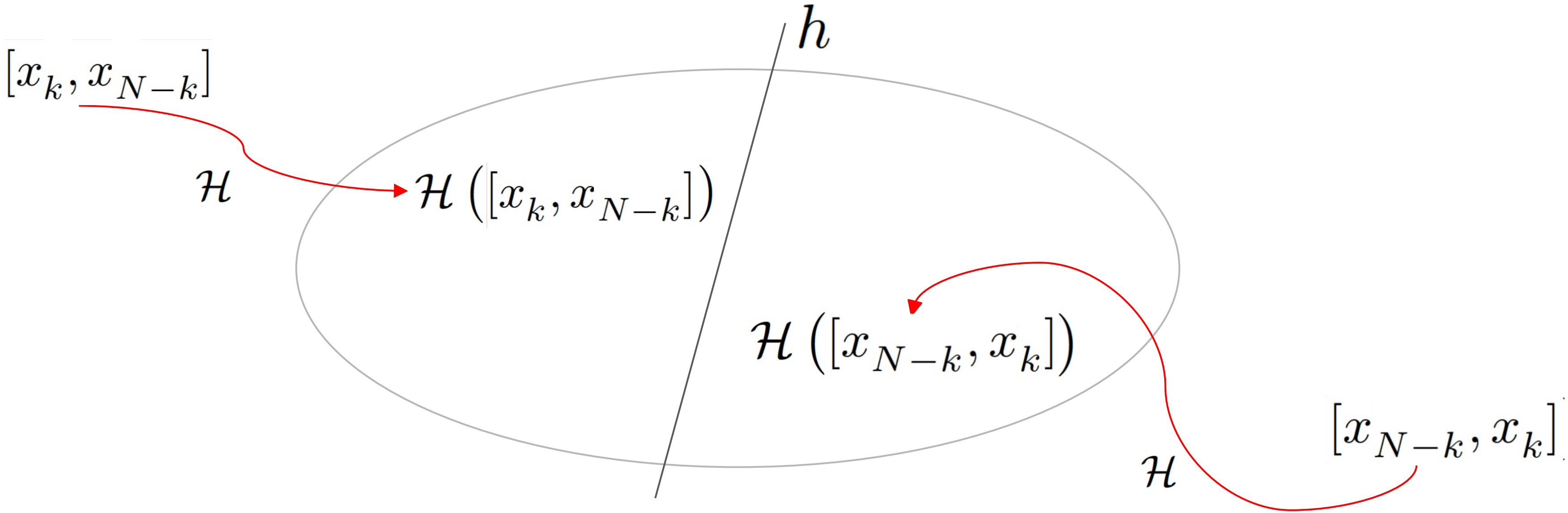
$[x_{N-k}, x_k]$

Proposed Method



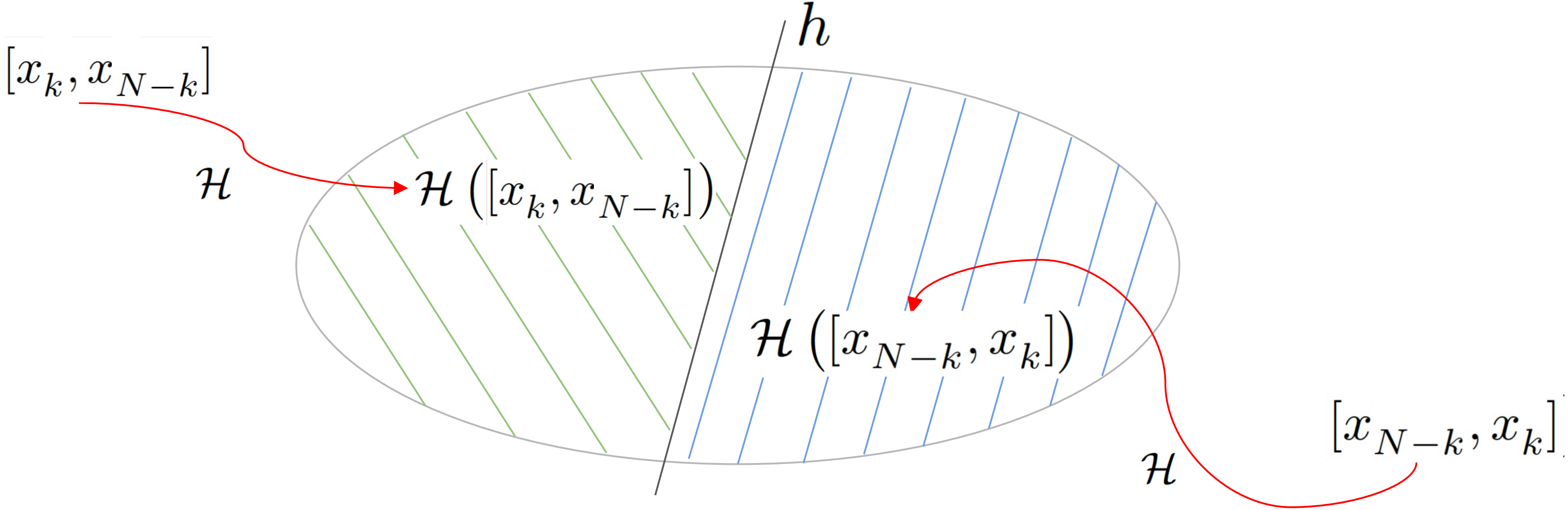
Proposed Method

JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA



h : hyperplane, $\underbrace{\mathcal{H} : \mathbb{R}^{2\times H\times W\times C} \rightarrow \mathbb{R}^d}_{\text{map each frame pair to an ordering scalar}}$

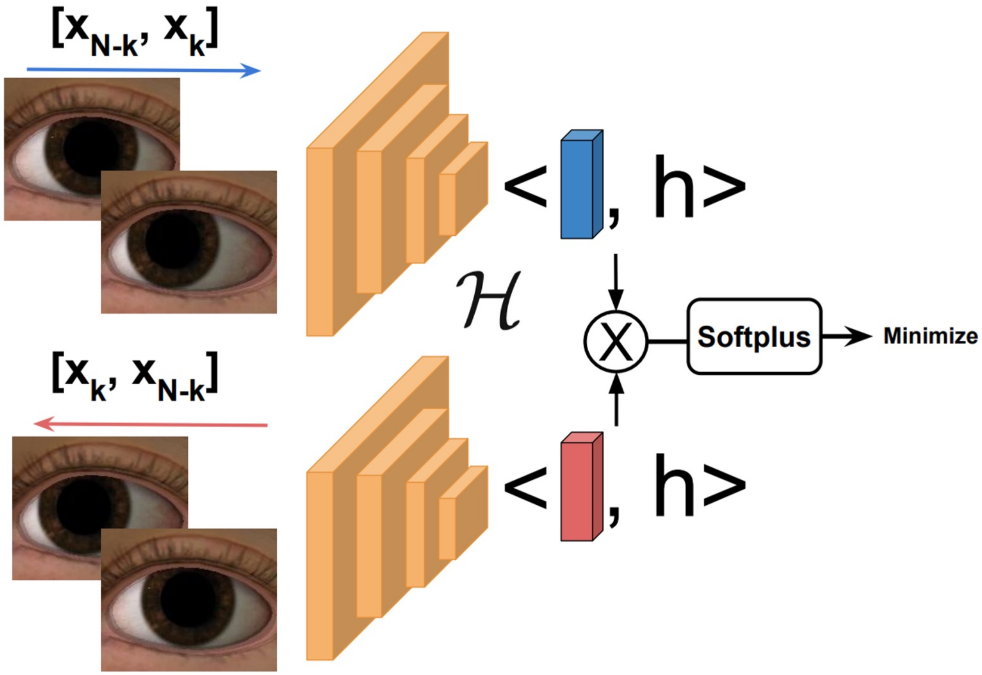
Proposed Method



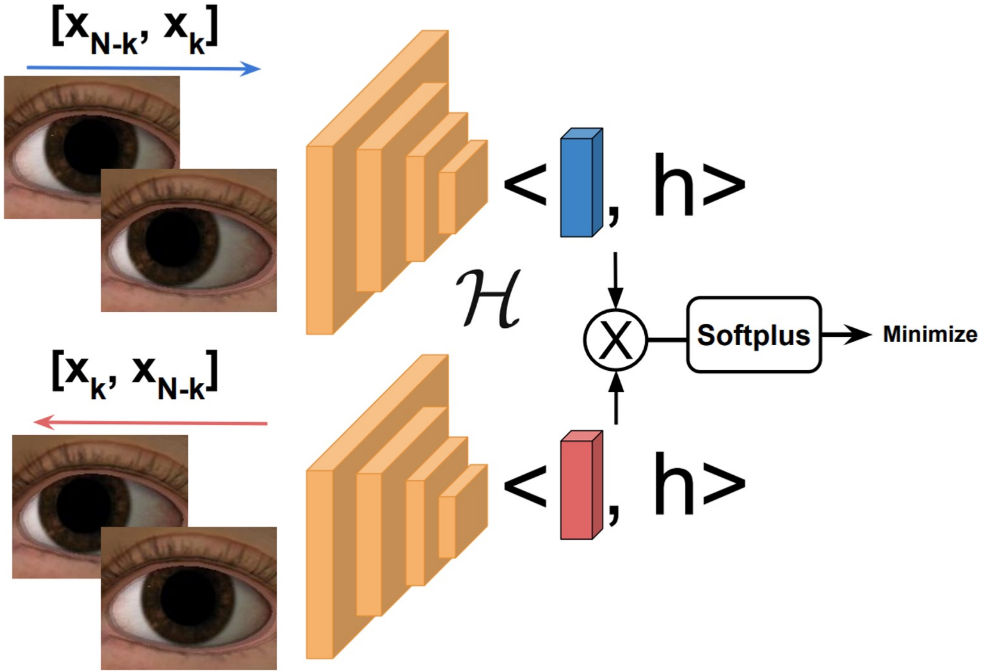
h : hyperplane, $\underbrace{\mathcal{H} : \mathbb{R}^{2\times H\times W\times C} \rightarrow \mathbb{R}^d}_{\text{map each frame pair to an ordering scalar}}$

$$\left\langle \mathcal{H}\left(\left[x_k^i, x_{N-k}^i\right]\right), h\right\rangle \left\langle \mathcal{H}\left(\left[x_{N-k}^i, x_k^i\right]\right), h\right\rangle < 0$$

Proposed Method - HyperCUT Training

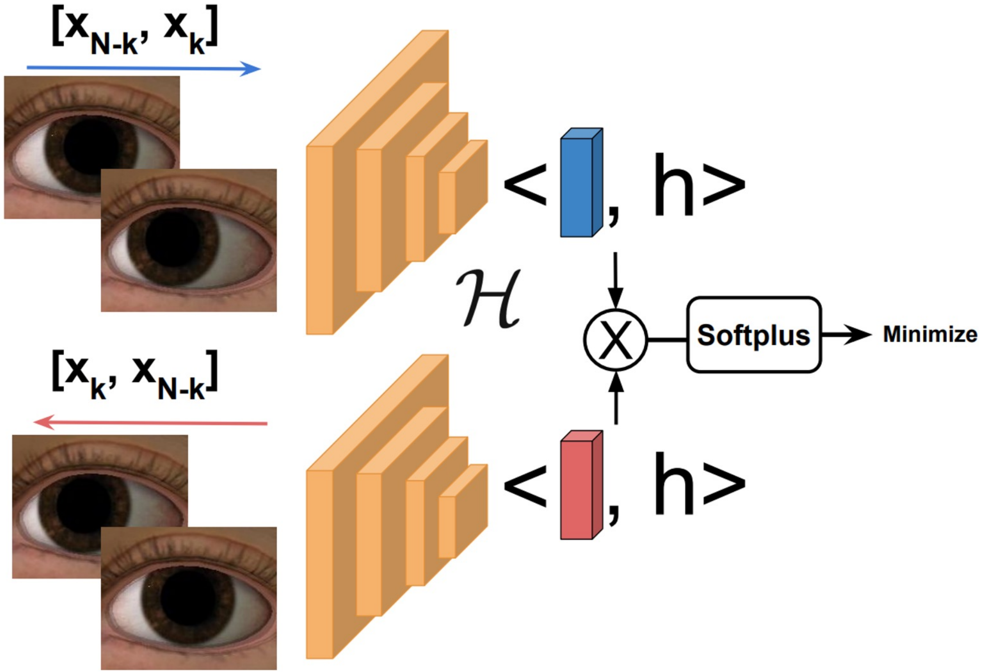


Proposed Method - HyperCUT Training



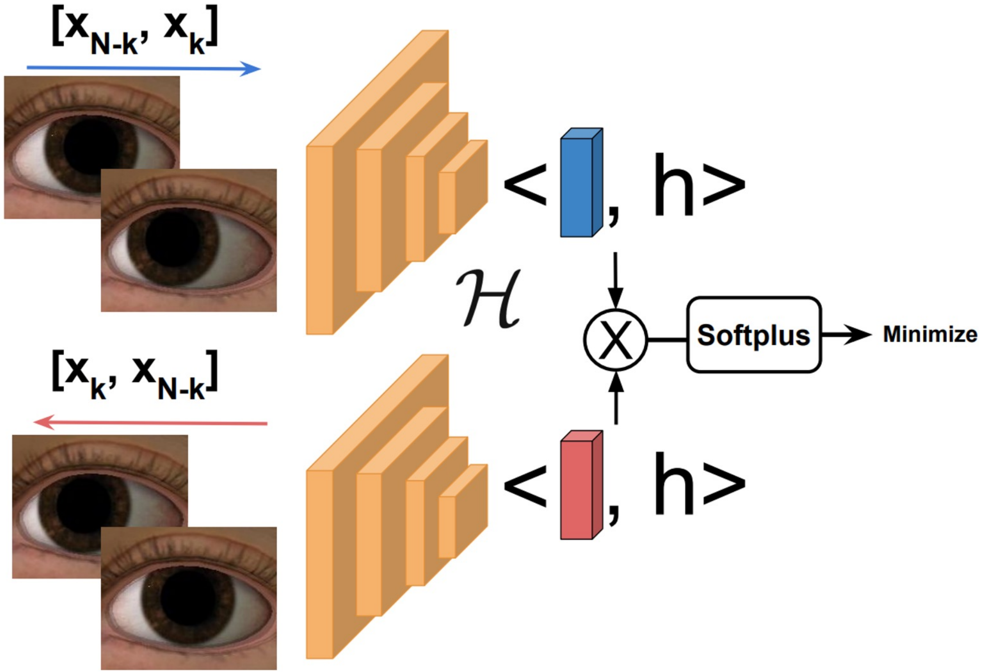
Proposed Method - HyperCUT Training

JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA



- For each sample i, our objective:

Proposed Method - HyperCUT Training

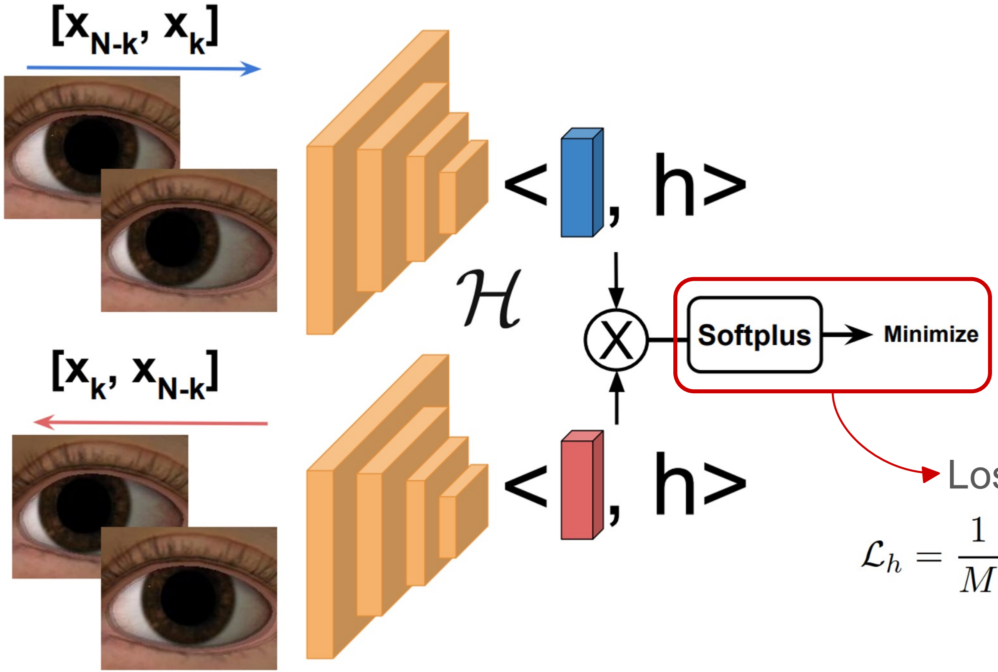


- For each sample i, our objective:

$$\mathcal{H}^* = \underset{\mathcal{H}}{\operatorname{argmin}}\, \sigma(\langle \mathcal{H}\left([x_k^i, x_{N-k}^i]\right), h\rangle \langle \mathcal{H}\left([x_{N-k}^i, x_k^i]\right), h\rangle)$$

$$\text{with } \sigma(z) = \mathit{Softmax}(z)$$

Proposed Method - HyperCUT Training



- For each sample i, our objective:

$$\mathcal{H}^* = \underset{\mathcal{H}}{\operatorname{argmin}}\, \sigma(\langle \mathcal{H}\left([x_k^i, x_{N-k}^i]\right), h\rangle \langle \mathcal{H}\left([x_{N-k}^i, x_k^i]\right), h\rangle)$$

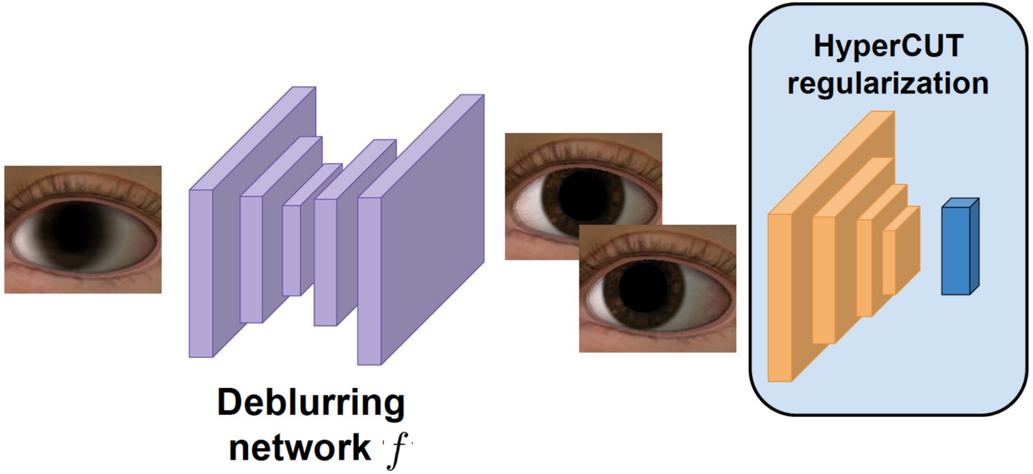
$$\text{with } \sigma(z) = \mathit{Softmax}(z)$$

Loss function:

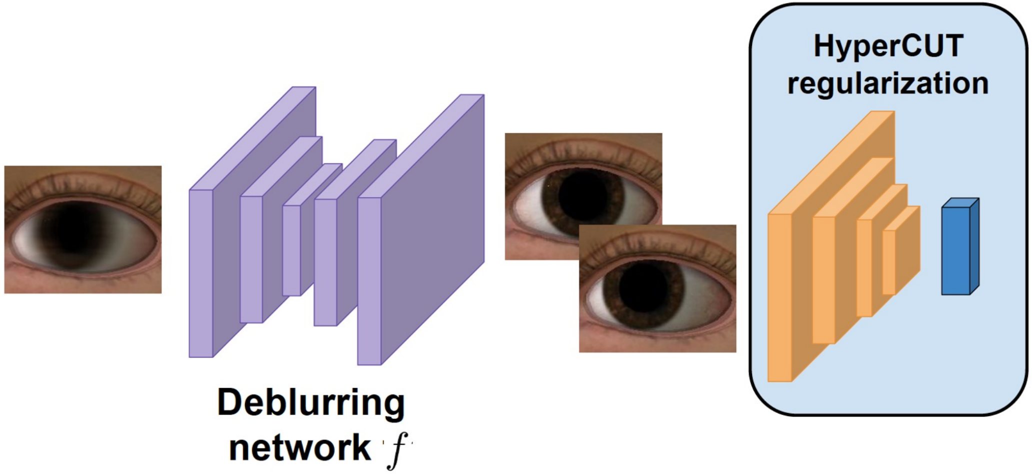
$$\mathcal{L}_h = \frac{1}{M}\sum_{i=1}^{M} softplus(\langle \mathcal{H}\left([x_k^i, x_{N-k}^i]\right), h\rangle\langle \mathcal{H}\left([x_{N-k}^i, x_k^i]\right), h\rangle)$$

With $softplus(t) = \log(1 + e^x)$

HyperCUT Regularization



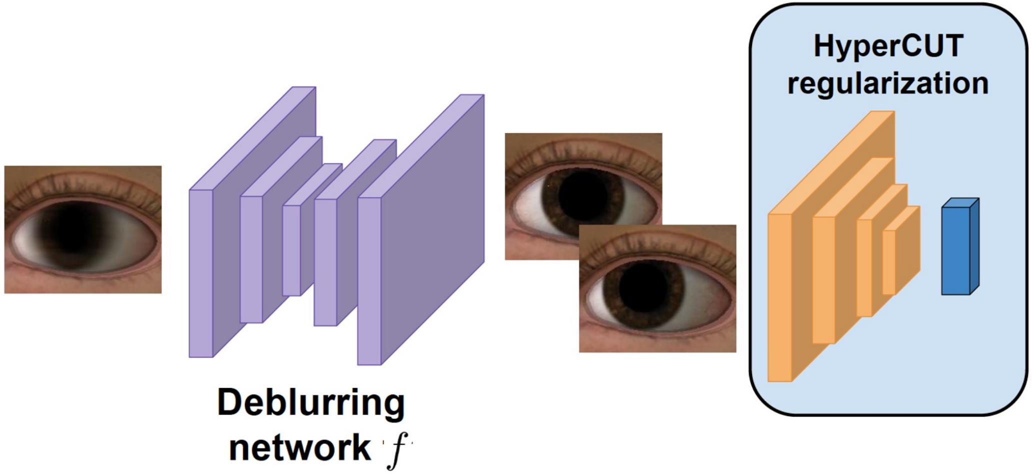
HyperCUT Regularization



$$\mathcal{R}_{hyp}(f) = \frac{1}{M}\sum_{i=1}^{M}\sum_{k=0}^{\lfloor N/2 \rfloor}\left\langle \mathcal{H}\left(f_k(y^i), f_{N-k}(y^i)\right), h\right\rangle$$

force to be on one side of h

HyperCUT Regularization



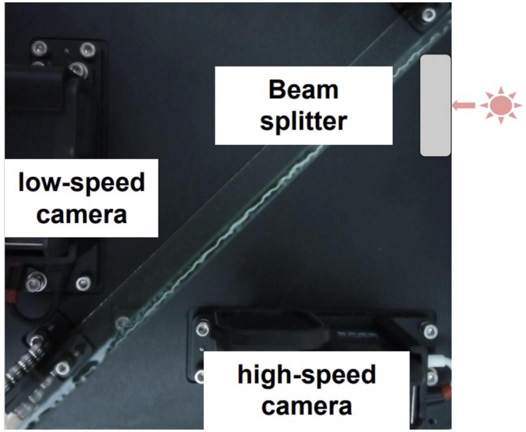
$$\mathcal{R}_{hyp}(f) = \frac{1}{M}\sum_{i=1}^{M}\sum_{k=0}^{\lfloor N/2 \rfloor}\left\langle \mathcal{H}\left(f_k(y^i), f_{N-k}(y^i)\right), h\right\rangle$$

Total Deblurring Loss: $\mathcal{L}(f) = \frac{1}{M}\sum_{i=1}^{M}\mathcal{L}_D(f(y^i)) + \alpha\mathcal{R}_{hyp}(f)$ where $\mathcal{L}_D$ can be any deblurring loss

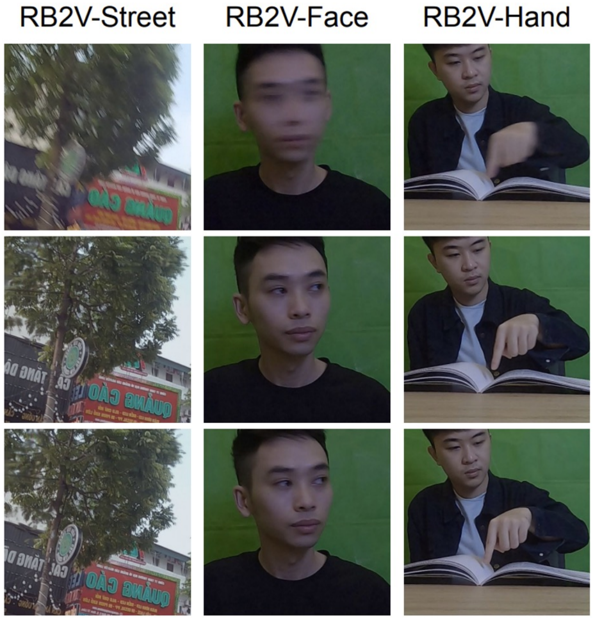
Datasets

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

Datasets - Real



Data subset	#data samples	
	Train	Test
RB2V-Street	9000	2053
RB2V-Face	8000	2157
RB2V-Hand	12000	4722



- We propose a real-world blur2vid dataset
- Three categories: Street, Face, and Hand

Datasets - Synthetic

- REDS
  - 120fps
  - Interpolate frames to form 7 sharp sequences
- B-Aist++
  - Config augmentation and setting proposed as in [3]

[3] *Zhihang Zhong, Xiao Sun, Zhirong Wu, Yinqiang Zheng, Stephen Lin, and Imari Sato. Animation from blur: Multimodal blur decomposition with motion guidance. ECCV, 2022*

Results

Results - Order Accuracy of HyperCUT

- We define 2 metrics:
  - ***hit***: is the ratio of frame pairs (x_k, x_{N-k}) that satisfy:

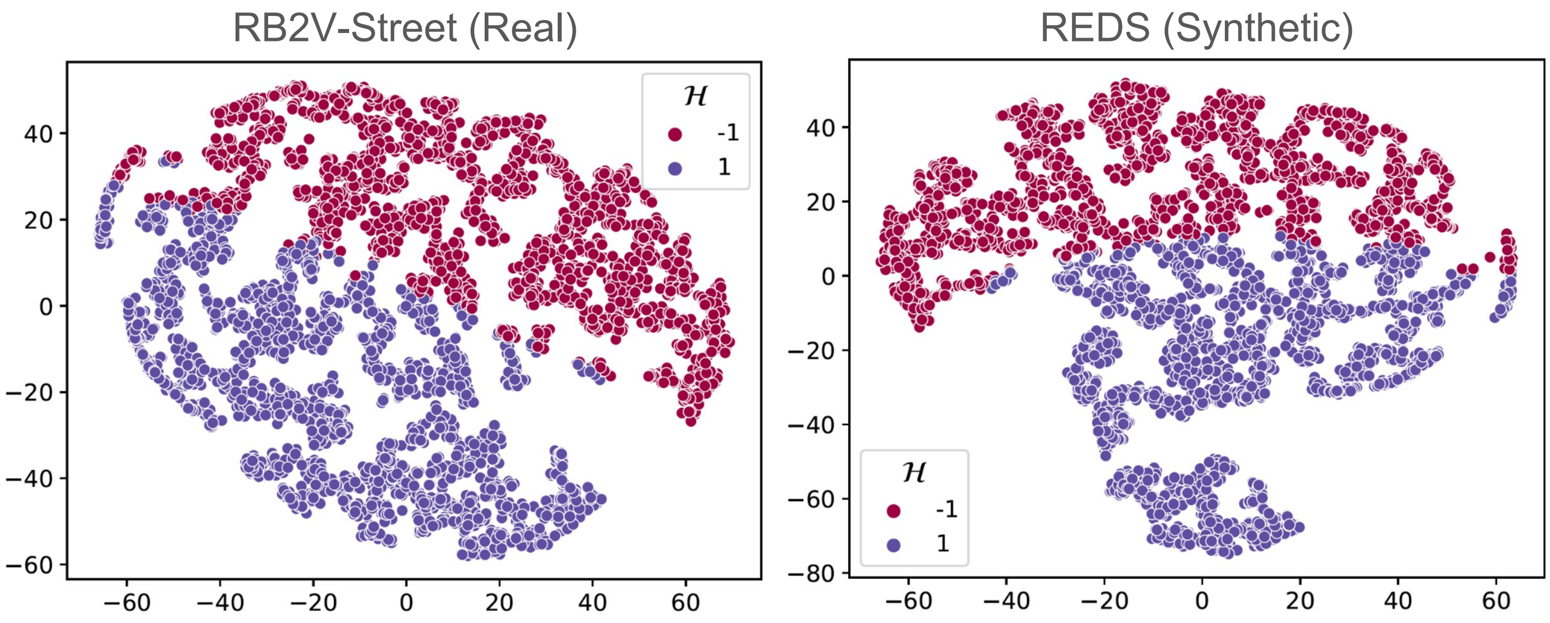
$$\langle \mathcal{H}\left(\left[x_k, x_{N-k}\right]\right), h\rangle \langle \mathcal{H}\left(\left[x_{N-k}, x_k\right]\right), h\rangle < 0$$

  - ***con***: measures the **con**sistency ratio that the pairs (x_1, x_7), (x_2, x_6), and (x_3, x_5) are in the same side of the hyperplane h.

Dataset	hit	$con@2$	$con@3$
REDS	95.7	96.5	94.4
B-Aist++	97.5	95.6	91.2
RB2V-Face	94.4	96.7	92.1
RB2V-Hand	98.6	97.0	96.7
RB2V-Street	98.7	98.3	96.8

Results - Order Accuracy of HyperCUT

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA



The t-SNE visualization of HyperCUT ordering mapping

Results - HyperCUT Effectiveness

- For fair comparison, as forward and backward frames are plausible solutions, we define metric pM where M is represented for **PSNR, SSIM, LPIPS**

$$\mathrm{pM}\left(\left[x_k^i, x_{N-k}^i\right]\right) = \max\left(\mathrm{M}\left(\left[x_k^i, x_{N-k}^i\right]\right), \mathrm{M}\left(\left[x_{N-k}^i, x_k^i\right]\right)\right)$$

[1] Meiguang Jin, Givi Meishvili, and Paolo Favaro. Learning to extract a video sequence from a single motion-blurred image. CVPR, 2018
[2] Kuldeep Purohit, Anshul Shah, and AN Rajagopalan. Bringing alive blurred moments. CVPR, 2019

Results - HyperCUT Effectiveness

Generated frames

Baselines

Model	Dataset	1^{st}	2^{nd}	3^{rd}	4^{th}	5^{th}	6^{th}	7^{th}
[1]	REDS	20.65	22.63	24.20	23.50	24.20	22.63	20.65
[1] + HyperCUT	REDS	**22.87**	**24.88**	**26.29**	**25.10**	**26.29**	**24.88**	**22.86**
[2]	REDS	22.78	24.47	26.14	**31.50**	26.12	24.49	22.83
[2] + HyperCUT	REDS	**26.75**	**28.30**	**29.42**	29.97	**29.41**	**28.30**	**26.76**
[2]	RB2V	26.99	27.99	29.45	**32.08**	29.55	28.06	27.04
[2] + HyperCUT	RB2V	**28.29**	**29.20**	**30.43**	**32.08**	**30.53**	**29.22**	**28.25**

The pPSNR↑ scores (dB)

[1] Meiguang Jin, Givi Meishvili, and Paolo Favaro. Learning to extract a video sequence from a single motion-blurred image. CVPR, 2018
[2] Kuldeep Purohit, Anshul Shah, and AN Rajagopalan. Bringing alive blurred moments. CVPR, 2019

Results - HyperCUT Effectiveness

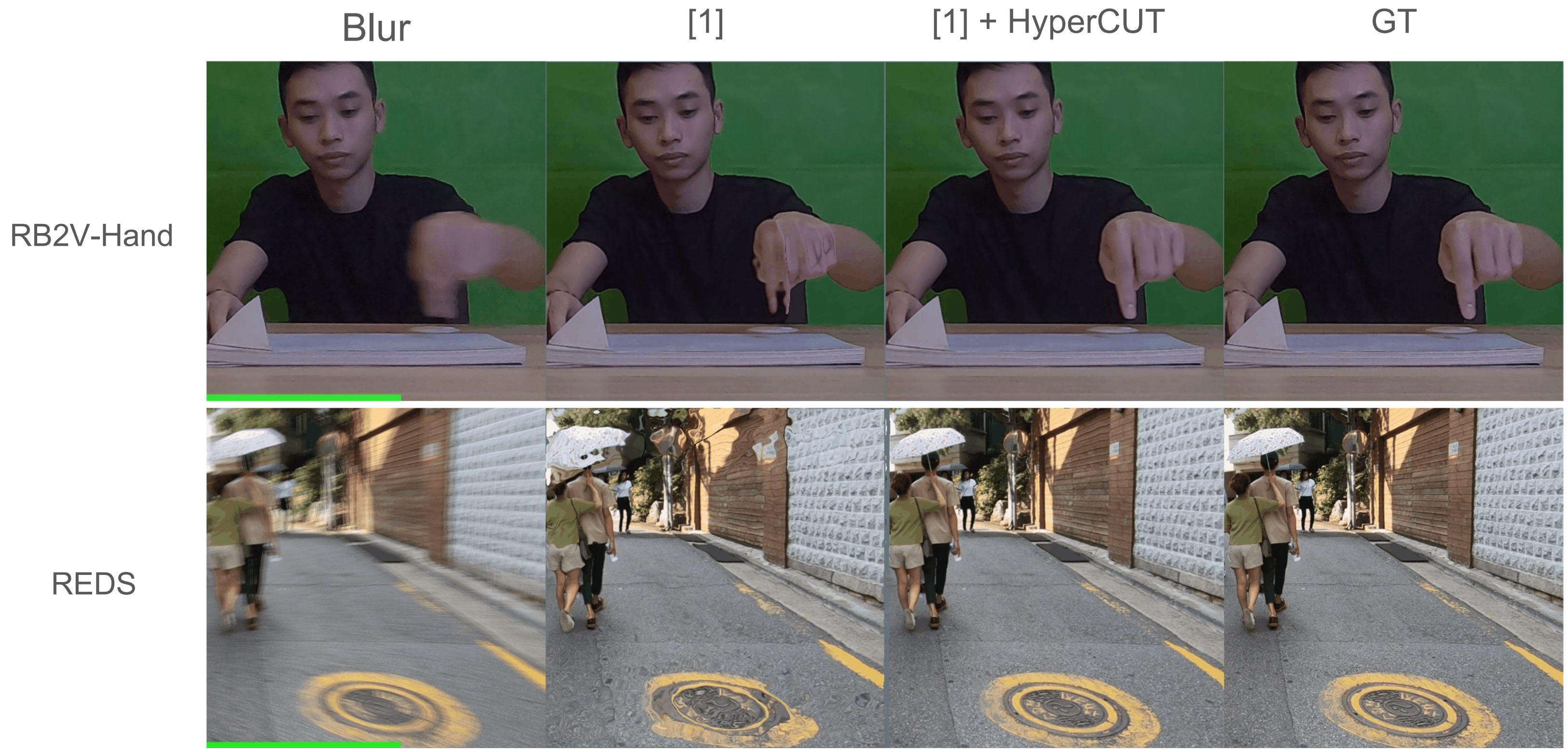
Baselines

Model		Generated frames						
		1^{st}	2^{nd}	3^{rd}	4^{th}	5^{th}	6^{th}	7^{th}
[1]	REDS	20.65	22.63	24.20	23.50	24.20	22.63	20.65
[1] + HyperCUT	REDS	**22.87**	**24.88**	**26.29**	**25.10**	**26.29**	**24.88**	**22.86**
[2]	REDS	22.78	24.47	26.14	**31.50**	26.12	24.49	22.83
[2] + HyperCUT	REDS	**26.75**	**28.30**	**29.42**	29.97	**29.41**	**28.30**	**26.76**
[2]	RB2V	26.99	27.99	29.45	**32.08**	29.55	28.06	27.04
[2] + HyperCUT	RB2V	**28.29**	**29.20**	**30.43**	**32.08**	**30.53**	**29.22**	**28.25**

The pPSNR↑ scores (dB)

[1] Meiguang Jin, Givi Meishvili, and Paolo Favaro. Learning to extract a video sequence from a single motion-blurred image. CVPR, 2018
[2] Kuldeep Purohit, Anshul Shah, and AN Rajagopalan. Bringing alive blurred moments. CVPR, 2019

Results - HyperCUT Effectiveness



[1] Meiguang Jin, Givi Meishvili, and Paolo Favaro. Learning to extract a video sequence from a single motion-blurred image. CVPR, 2018
[2] Kuldeep Purohit, Anshul Shah, and AN Rajagopalan. Bringing alive blurred moments. CVPR, 2019

Results - HyperCUT Effectiveness

Blur | [2] | [2] + HyperCUT | GT

RB2V-Street

REDS

[1] Meiguang Jin, Givi Meishvili, and Paolo Favaro. Learning to extract a video sequence from a single motion-blurred image. CVPR, 2018
[2] Kuldeep Purohit, Anshul Shah, and AN Rajagopalan. Bringing alive blurred moments. CVPR, 2019

Results - HyperCUT Effectiveness

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

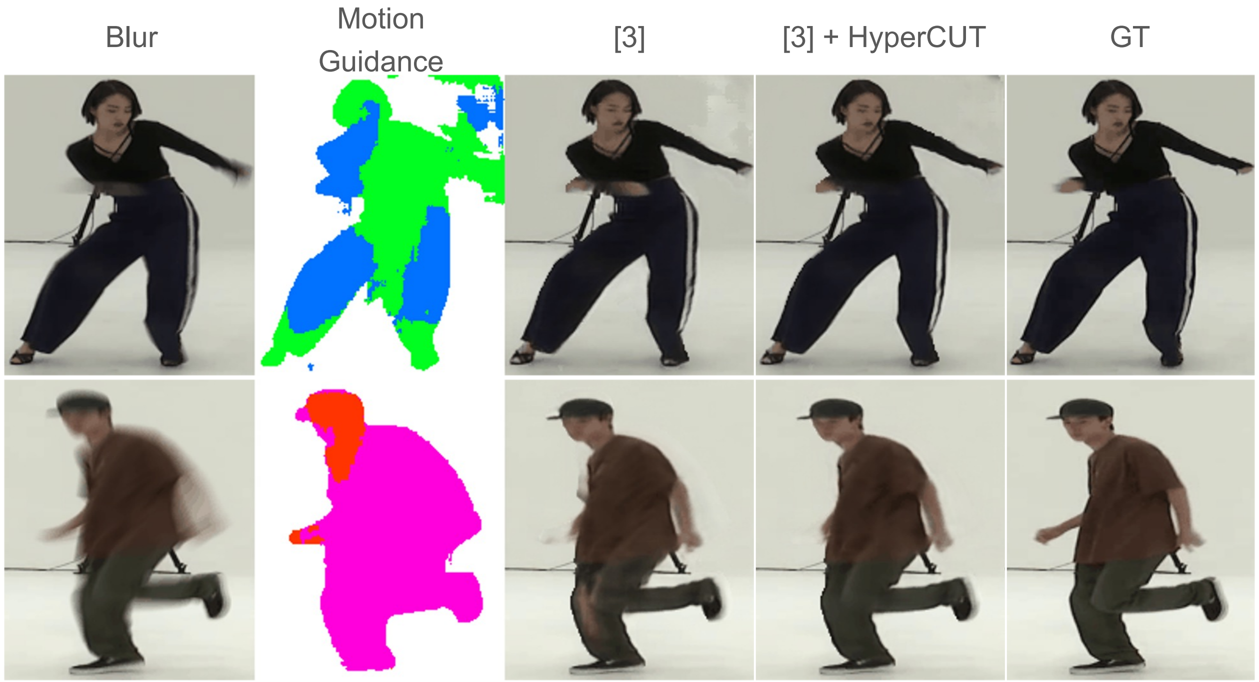
Results on the B-Aist++ dataset

Method	[3] (from paper)	[3] (reproduced)	[3] + HyperCUT
$\mathcal{P}_1$	19.97 / 0.860 / 0.089	20.58 / 0.890 / 0.068	22.16 / 0.901 / 0.102
$\mathcal{P}_3$	22.44 / 0.890 / 0.068	21.21 / 0.899 / 0.063	23.31 / 0.915 / 0.062
$\mathcal{P}_5$	23.49 / 0.911 / 0.060	22.48 / 0.903 / 0.061	23.81 / 0.920 / 0.060

The pPSNR↑/ pSSIM↑ / pLPIPS↓ scores

[3] Zhihang Zhong, Xiao Sun, Zhirong Wu, Yinqiang Zheng, Stephen Lin, and Imari Sato. Animation from blur: Multimodal blur decomposition with motion guidance. ECCV, 2022

Results - HyperCUT Effectiveness



[3] *Zhihang Zhong, Xiao Sun, Zhirong Wu, Yinqiang Zheng, Stephen Lin, and Imari Sato. Animation from blur: Multimodal blur decomposition with motion guidance. ECCV, 2022*

Conclusion

- We introduce **HyperCUT** which is used to solve the **order-ambiguity issue** effectively for the task of extracting a sharp video sequence from a blurry image.
- We build a new dataset for the task - **RB2V**, covering three categories: **Street**, **Face**, and **Hand**. This is the **first** real and large-scale dataset for image-to-video deblurring.
- Our model achieves **state-of-the-art performance** on both synthetic and real-world benchmarks.
- Future research on adapting HyperCUT for handling complex movements and long exposure blur would be an interesting avenue for exploration.

Code

Paper