Learning Spatially Regularized Correlation Filters for Visual Tracking

Martin Danelljan, Gustav Häger,
Fahad Khan, Michael Felsberg

Computer Vision Laboratory, Linköping University, Sweden

SRDCF

- **ICCV 2015** paper
  - –Poster session 4B on Wednesday afternoon
- **Won** the "OpenCV State of The Art Vision Challenge in Tracking"
- **Best** results in ICCV 2015 on OTB-2013
- **Competitive** results on VOT2015 and VOT-TIR2015
- Webpage and Matlab code

Discriminative Correlation Filters (DCF)

Standard DCF Formulation

$$S_f(x) = \sum_{l=1}^{d} x^l * f^l$$

Standard DCF Formulation

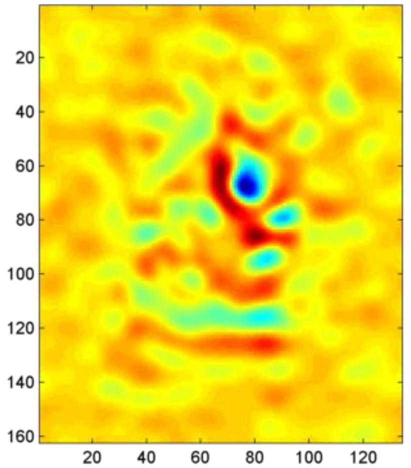
$$S_f(x) = \sum_{l=1}^{d} x^l * f^l$$

Standard DCF Formulation

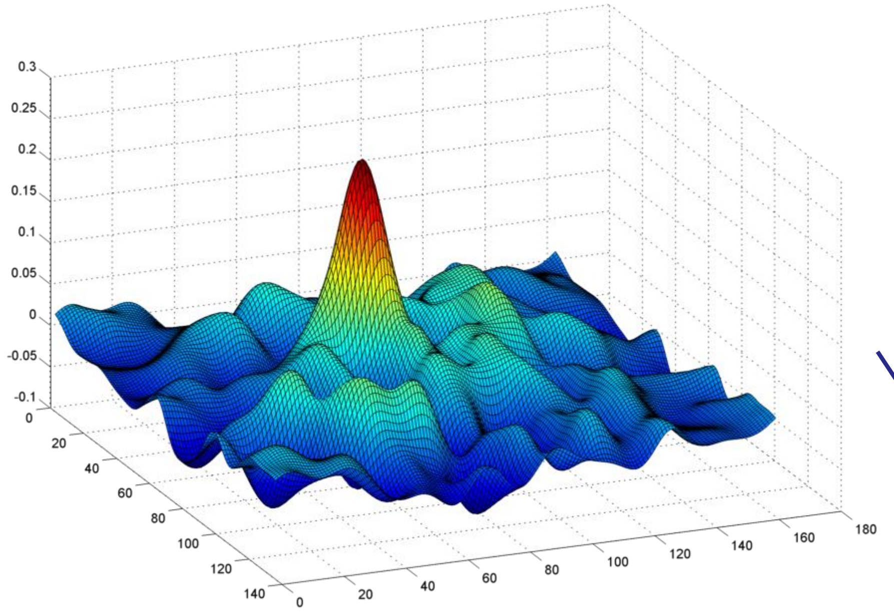
$$S_f(x) = \sum_{l=1}^{d} x^l * f^l$$

Standard DCF Formulation

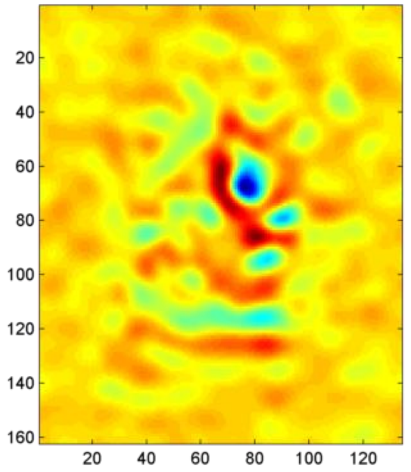
$$S_f(x) = \sum_{l=1}^{d} x^l * f^l$$



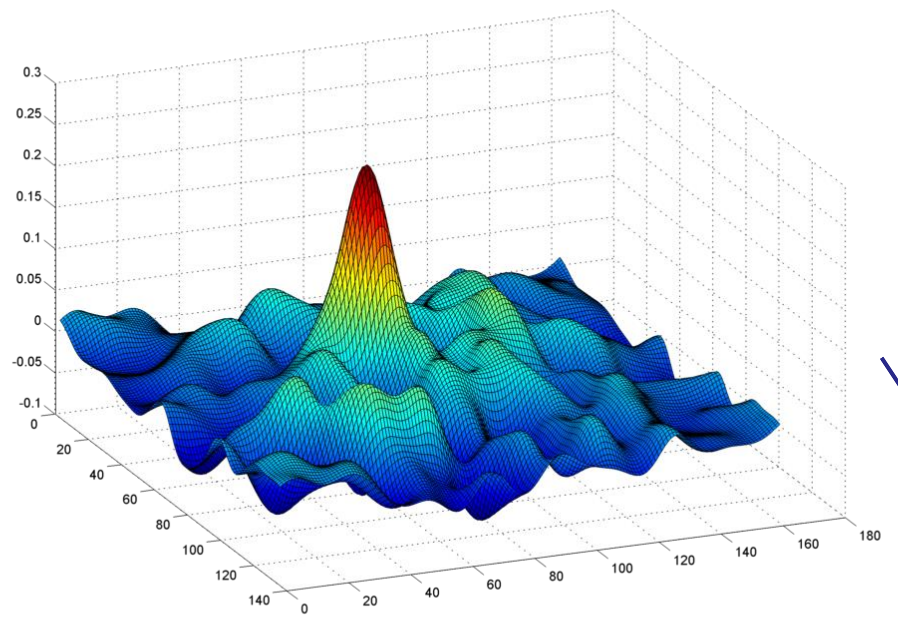
Standard DCF Formulation

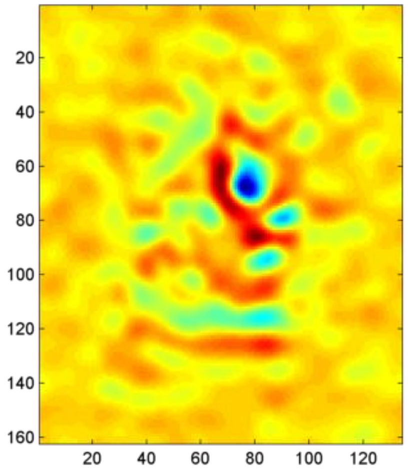


$$S_f(x) = \sum_{l=1}^{d} x^l * f^l$$



Standard DCF Formulation

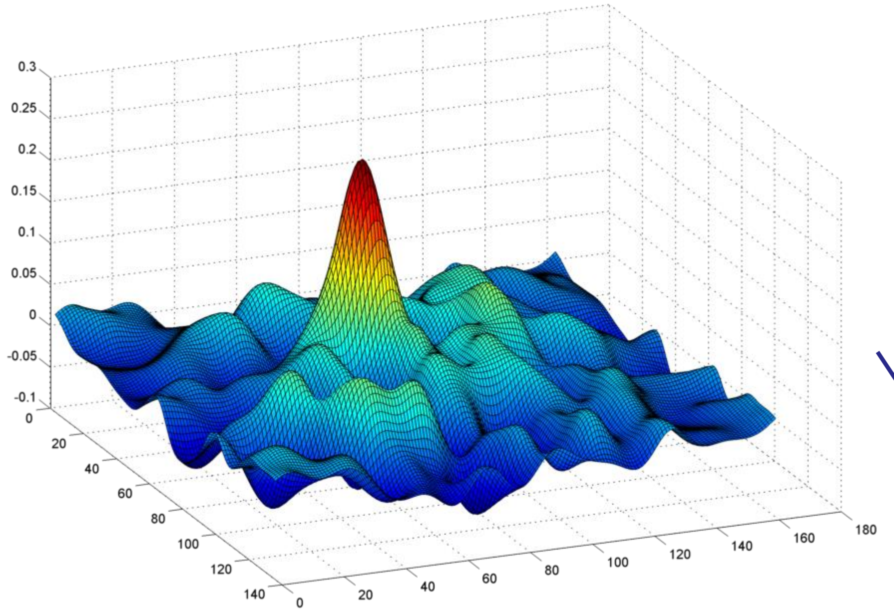




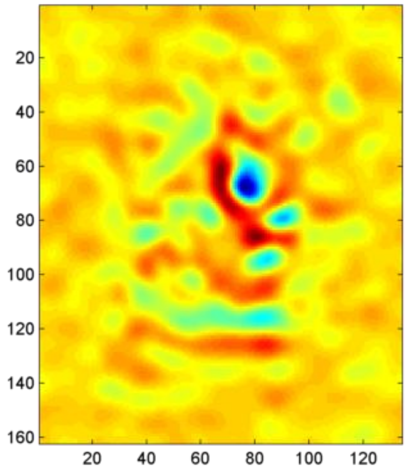
$$S_f(x) = \sum_{l=1}^{d} x^l * f^l$$

$$\varepsilon_t(f) = \sum_{k=1}^{t} \alpha_k \left\| S_f(x_k) - y_k \right\|^2 + \lambda \sum_{l=1}^{d} \left\| f^l \right\|^2$$

Standard DCF Formulation

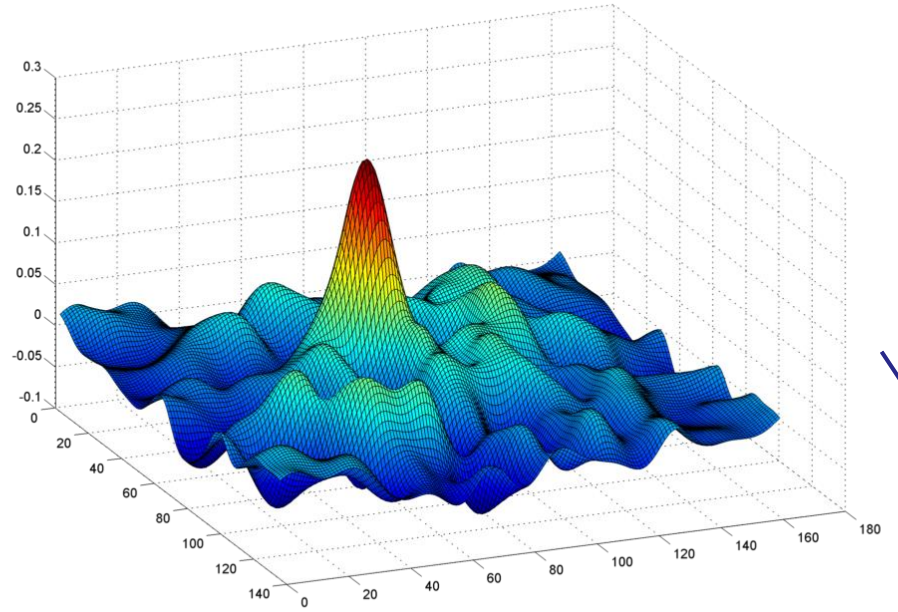


$$S_f(x) = \sum_{l=1}^{d} x^l * f^l$$

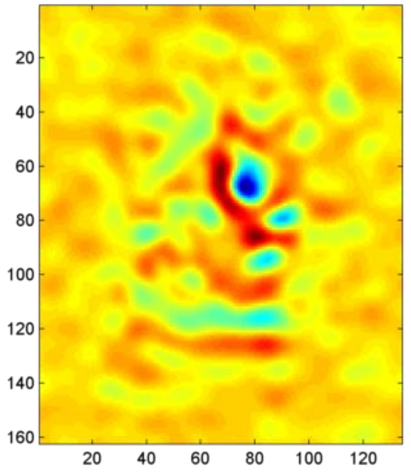


$$\varepsilon_t(f) = \sum_{k=1}^{t} \alpha_k \left\| S_f(x_k) - y_k \right\|^2 + \lambda \sum_{l=1}^{d} \left\| f^l \right\|^2$$

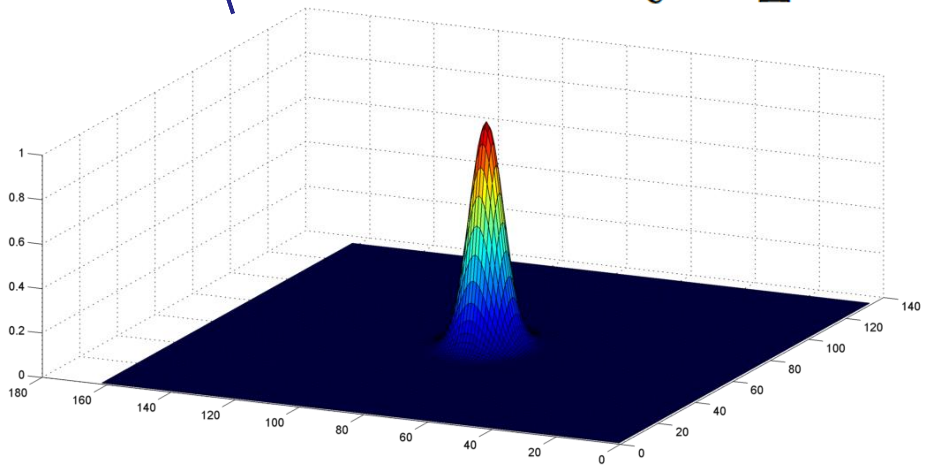
Standard DCF Formulation



$$S_f(x) = \sum_{l=1}^{d} x^l * f^l$$



$$\varepsilon_t(f) = \sum_{k=1}^{t} \alpha_k \left\| S_f(x_k) - y_k \right\|^2 + \lambda \sum_{l=1}^{d} \left\| f^l \right\|^2$$



Standard DCF Formulation

- Linear least squares problem
- Diagonalizable by the DFT if $d = 1$
- Incremental update schemes
  - –Henriques et al. (ECCV 2012)
  - –Danelljan et al. (CVPR 2014)
  - –Danelljan et al. (BMVC 2014)
- Based an **harsh approximations**

Standard DCF Formulation

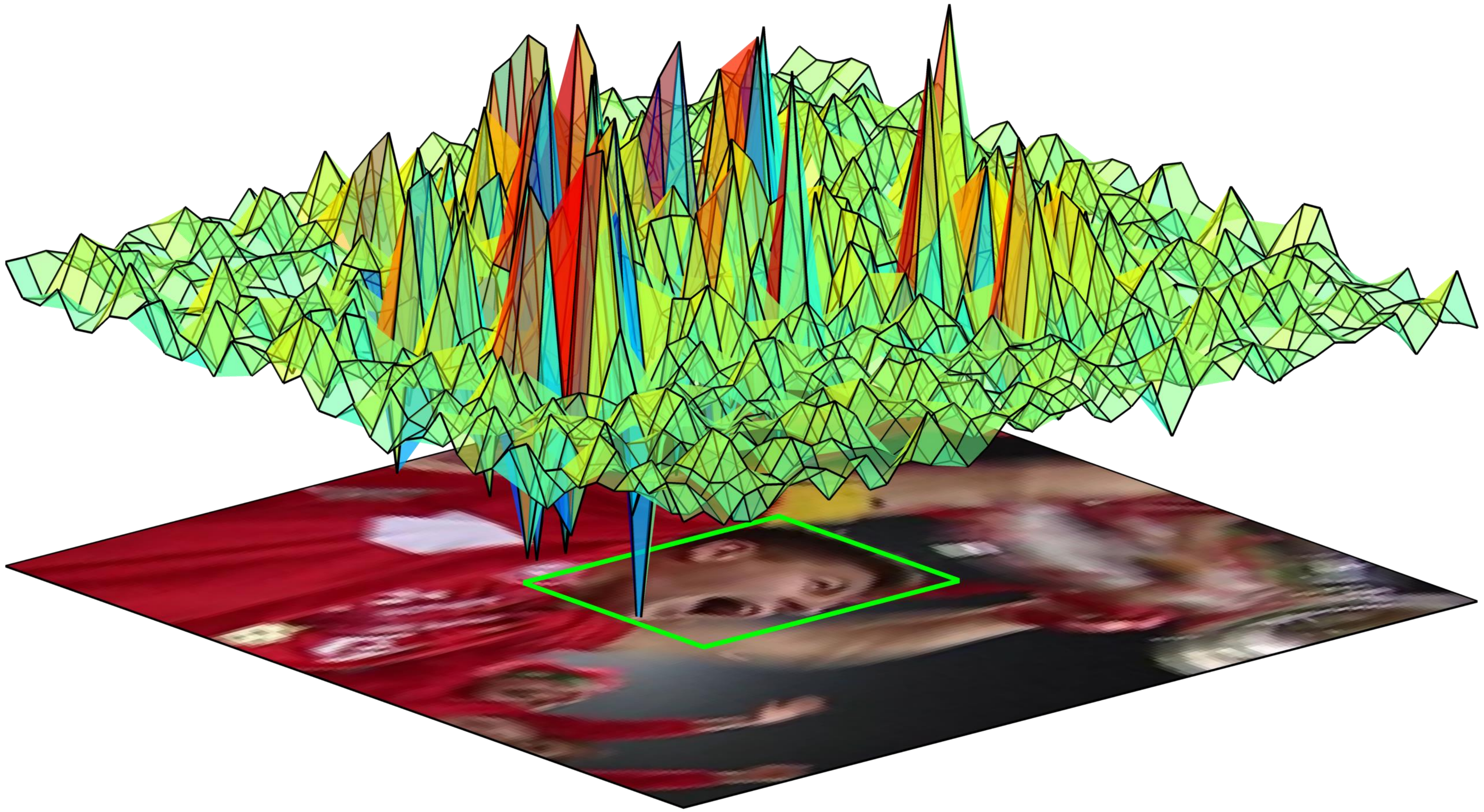
…has a major flaw…

Circular Convolution $\Longleftrightarrow$ Periodic Extension

Larger Samples?

Larger Samples?

Why?



Limitations of a Small Sample Size

Forces a **small** sample size in training/detection

- Limited search region
- Limited training data
- Corrupted samples

Previous Work

- Optimizing a constraint filter
  - Fernandez et al. (PAMI 2015)
  - Galoogahi et al. (CVPR 2015)
- Leads to iterative optimization
  - Transition between spatial and Fourier domain
- Our approach:
  - More flexible
  - Efficient optimization in the Fourier domain

Our Approach: SRDCF

$$\varepsilon_t(f) = \sum_{k=1}^{t} \alpha_k \left\| S_f(x_k) - y_k \right\|^2 + \lambda \sum_{l=1}^{d} \left\| f^l \right\|^2$$

Our Approach: SRDCF

$$\varepsilon_t(f) = \sum_{k=1}^{t} \alpha_k \left\| S_f(x_k) - y_k \right\|^2 + \lambda \sum_{l=1}^{d} \left\| f^l \right\|^2$$

Our Approach: SRDCF

$$\varepsilon_t(f) = \sum_{k=1}^{t} \alpha_k \left\| S_f(x_k) - y_k \right\|^2 + \lambda \sum_{l=1}^{d} \left\| f^l \right\|^2$$

$$\Downarrow$$

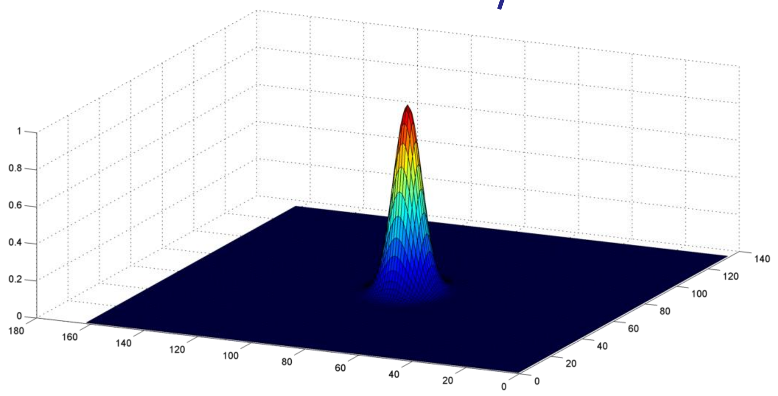
$$\varepsilon_t(f) = \sum_{k=1}^{t} \alpha_k \left\| S_f(x_k) - y_k \right\|^2 + \sum_{l=1}^{d} \left\| w \cdot f^l \right\|^2$$

Our Approach: SRDCF

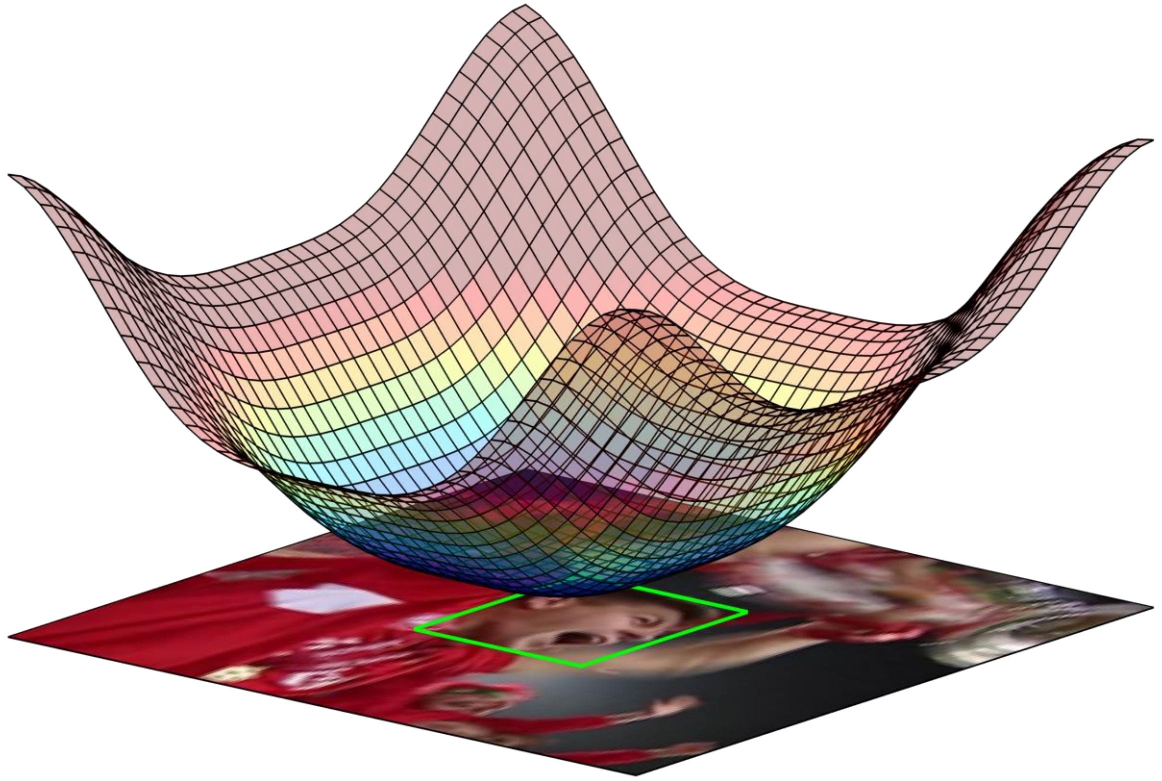
$$\varepsilon_t(f) = \sum_{k=1}^{t} \alpha_k \left\| S_f(x_k) - y_k \right\|^2 + \sum_{l=1}^{d} \left\| w \cdot f^l \right\|^2$$

Our Approach: SRDCF

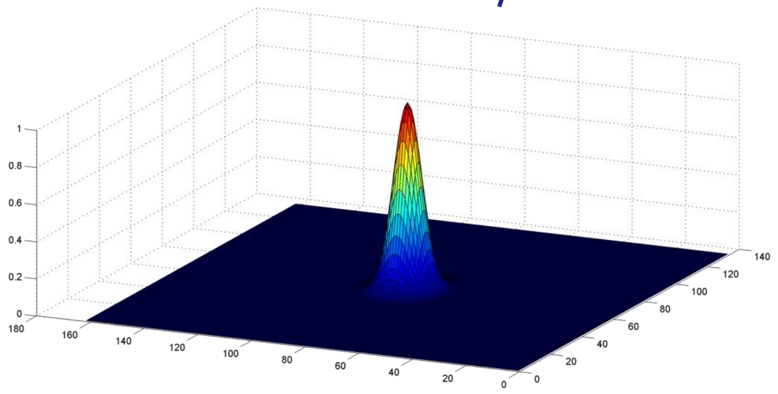
$$\varepsilon_t(f) = \sum_{k=1}^{t} \alpha_k \left\| S_f(x_k) - y_k \right\|^2 + \sum_{l=1}^{d} \left\| w \cdot f^l \right\|^2$$



Our Approach: SRDCF



$$\varepsilon_t(f) = \sum_{k=1}^{t} \alpha_k \left\| S_f(x_k) - y_k \right\|^2 + \sum_{l=1}^{d} \left\| w \cdot f^l \right\|^2$$



Our Approach: SRDCF

$$\varepsilon_t(f) = \sum_{k=1}^{t} \alpha_k \left\| S_f(x_k) - y_k \right\|^2 + \sum_{l=1}^{d} \left\| w \cdot f^l \right\|^2$$

Our Approach: SRDCF

DFT

$$\varepsilon_t(f) = \sum_{k=1}^{t} \alpha_k \left\| S_f(x_k) - y_k \right\|^2 + \sum_{l=1}^{d} \left\| w \cdot f^l \right\|^2$$

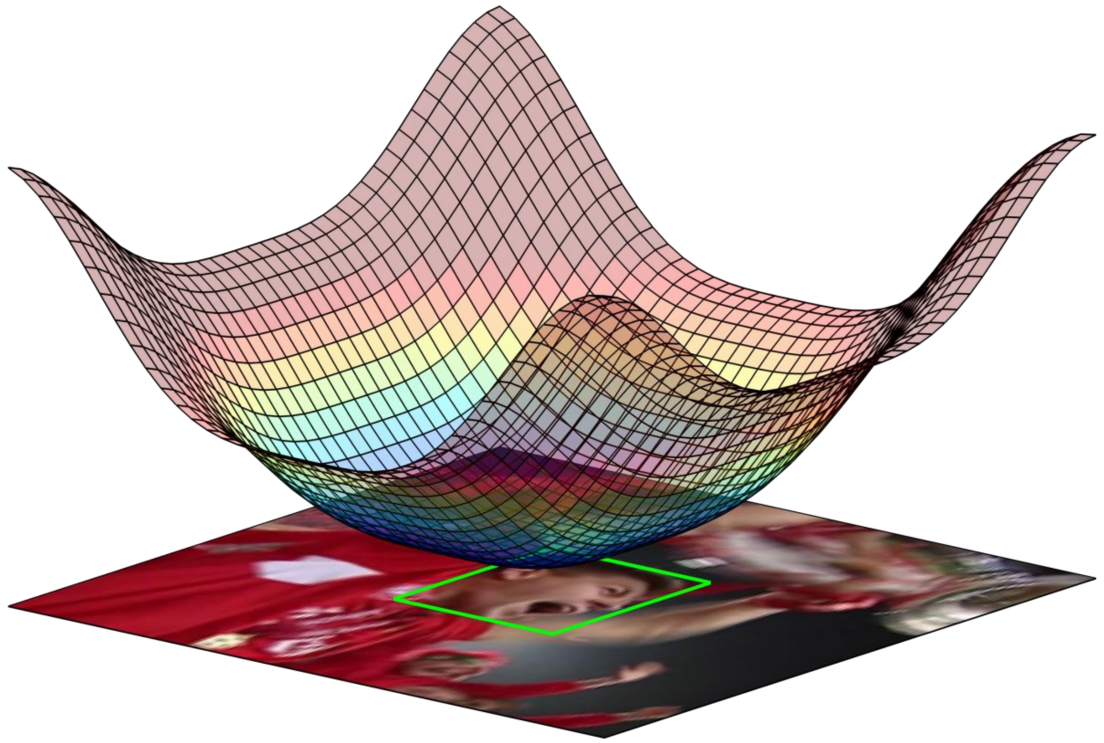
Our Approach: SRDCF

$$\check{\varepsilon}_t(\hat{f}) = \sum_{k=1}^{t} \alpha_k \left\| \sum_{l=1}^{d} \hat{x}_k^l \cdot \hat{f}^l - \hat{y}_k \right\|^2 + \sum_{l=1}^{d} \left\| \frac{\hat{w}}{MN} * \hat{f}^l \right\|^2$$

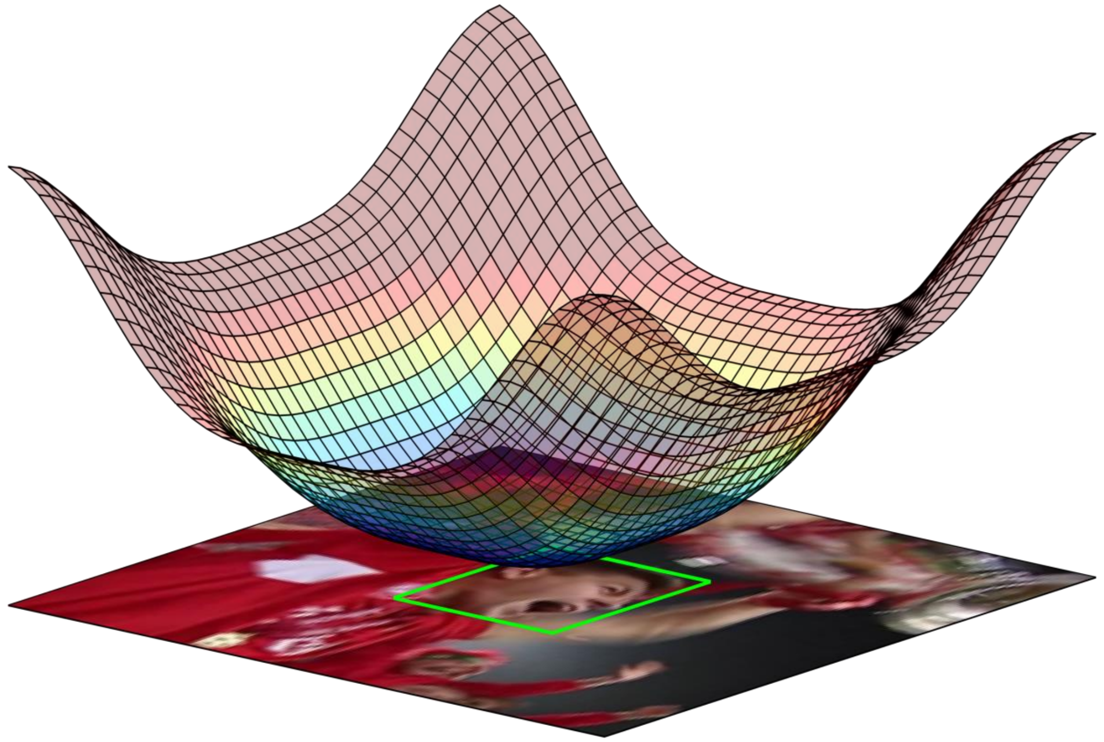
DFT $\Uparrow$

$$\varepsilon_t(f) = \sum_{k=1}^{t} \alpha_k \left\| S_f(x_k) - y_k \right\|^2 + \sum_{l=1}^{d} \left\| w \cdot f^l \right\|^2$$

Our Approach: SRDCF

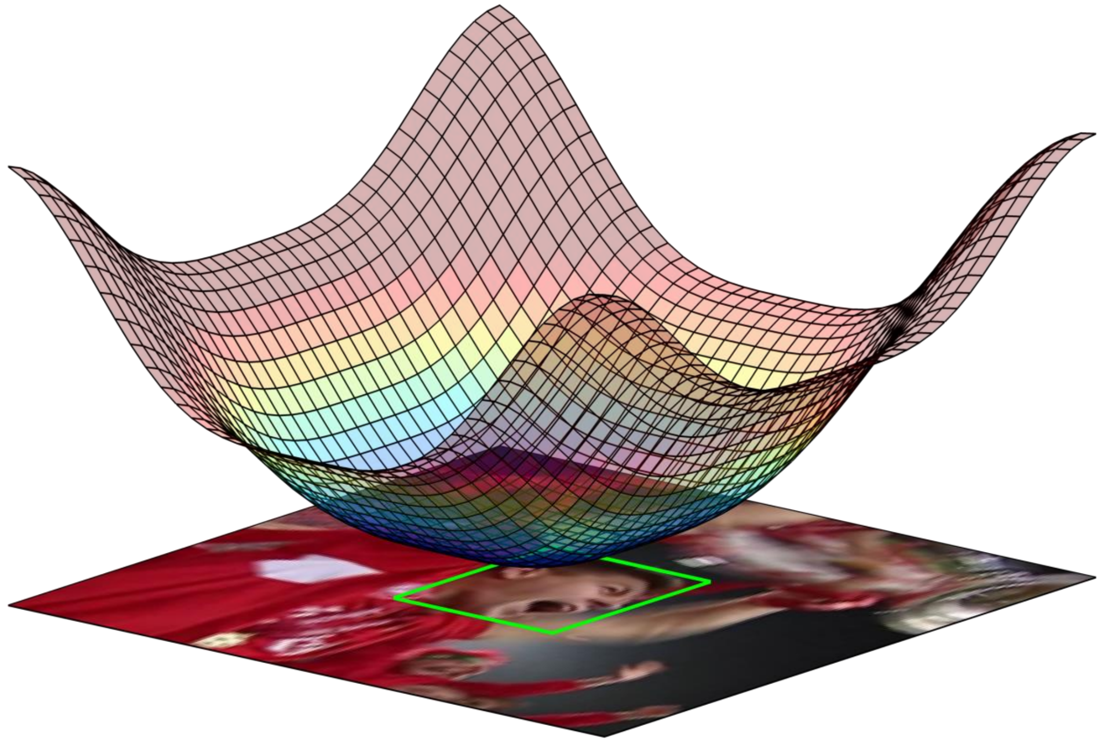


Our Approach: SRDCF

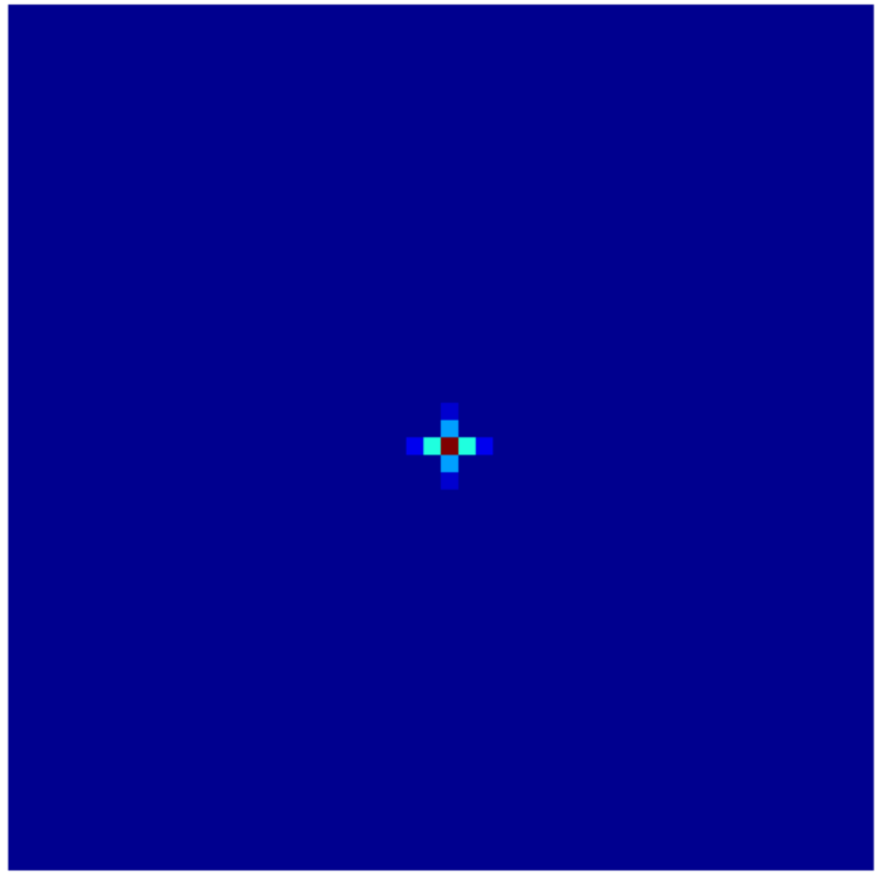


DFT $\longmapsto$

Our Approach: SRDCF



$\xrightarrow{\text{DFT}}$



Our Approach: SRDCF

$$\check{\varepsilon}_t(\hat{f}) = \sum_{k=1}^{t} \alpha_k \left\| \sum_{l=1}^{d} \hat{x}_k^l \cdot \hat{f}^l - \hat{y}_k \right\|^2 + \sum_{l=1}^{d} \left\| \frac{\hat{w}}{MN} * \hat{f}^l \right\|^2$$

Our Approach: SRDCF

$$\check{\varepsilon}_t(\hat{f}) = \sum_{k=1}^{t} \alpha_k \left\| \sum_{l=1}^{d} \hat{x}_k^l \cdot \hat{f}^l - \hat{y}_k \right\|^2 + \sum_{l=1}^{d} \left\| \frac{\hat{w}}{MN} * \hat{f}^l \right\|^2$$

$$\hat{\mathbf{x}}_k^l, \hat{\mathbf{f}}^l, \hat{\mathbf{y}} \in \mathbb{C}^{MN}$$

Our Approach: SRDCF

$$\check{\varepsilon}_t(\hat{f}) = \sum_{k=1}^{t} \alpha_k \left\| \sum_{l=1}^{d} \hat{x}_k^l \cdot \hat{f}^l - \hat{y}_k \right\|^2 + \sum_{l=1}^{d} \left\| \frac{\hat{w}}{MN} * \hat{f}^l \right\|^2$$

$$\Downarrow$$

$$\check{\varepsilon}_t(\hat{f}) = \sum_{k=1}^{t} \alpha_k \left\| \sum_{l=1}^{d} \mathcal{D}(\hat{\mathbf{x}}_k^l) \hat{\mathbf{f}}^l - \hat{\mathbf{y}}_k \right\|^2 + \sum_{l=1}^{d} \left\| \frac{\mathcal{C}(\hat{\mathbf{w}})}{MN} \hat{\mathbf{f}}^l \right\|^2.$$

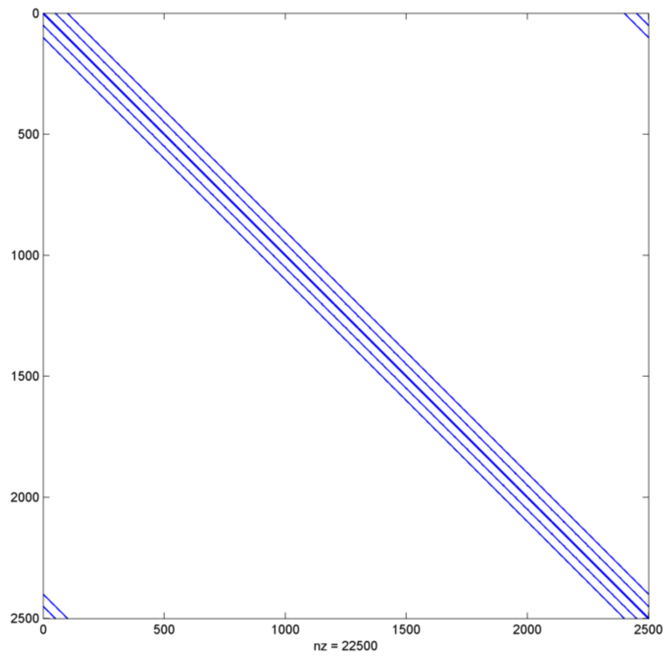
$$\hat{\mathbf{x}}_k^l, \hat{\mathbf{f}}^l, \hat{\mathbf{y}} \in \mathbb{C}^{MN}$$

Our Approach: SRDCF

$$\check{\varepsilon}_t(\hat{f}) = \sum_{k=1}^{t} \alpha_k \left\| \sum_{l=1}^{d} \hat{x}_k^l \cdot \hat{f}^l - \hat{y}_k \right\|^2 + \sum_{l=1}^{d} \left\| \frac{\hat{w}}{MN} * \hat{f}^l \right\|^2$$

$$\check{\varepsilon}_t(\hat{f}) = \sum_{k=1}^{t} \alpha_k \left\| \sum_{l=1}^{d} \mathcal{D}(\hat{\mathbf{x}}_k^l) \hat{\mathbf{f}}^l - \hat{\mathbf{y}}_k \right\|^2 + \sum_{l=1}^{d} \left\| \frac{\mathcal{C}(\hat{\mathbf{w}})}{MN} \hat{\mathbf{f}}^l \right\|^2.$$

$$\hat{\mathbf{x}}_k^l, \hat{\mathbf{f}}^l, \hat{\mathbf{y}} \in \mathbb{C}^{MN}$$



Our Approach: SRDCF

$$\check{\varepsilon}_t(\hat{f}) = \sum_{k=1}^{t} \alpha_k \left\| \sum_{l=1}^{d} \hat{x}_k^l \cdot \hat{f}^l - \hat{y}_k \right\|^2 + \sum_{l=1}^{d} \left\| \frac{\hat{w}}{MN} * \hat{f}^l \right\|^2$$

$$\check{\varepsilon}_t(\hat{f}) = \sum_{k=1}^{t} \alpha_k \left\| \sum_{l=1}^{d} \mathcal{D}(\hat{\mathbf{x}}_k^l)\hat{\mathbf{f}}^l - \hat{\mathbf{y}}_k \right\|^2 + \sum_{l=1}^{d} \left\| \frac{\mathcal{C}(\hat{\mathbf{w}})}{MN} \hat{\mathbf{f}}^l \right\|^2.$$

$$\hat{\mathbf{x}}_k^l, \hat{\mathbf{f}}^l, \hat{\mathbf{y}} \in \mathbb{C}^{MN}$$

Our Approach: SRDCF

$$\check{\varepsilon}_t(\hat{f}) = \sum_{k=1}^{t} \alpha_k \left\| \sum_{l=1}^{d} \hat{x}_k^l \cdot \hat{f}^l - \hat{y}_k \right\|^2 + \sum_{l=1}^{d} \left\| \frac{\hat{w}}{MN} * \hat{f}^l \right\|^2$$

$$\downarrow$$

$$\check{\varepsilon}_t(\hat{f}) = \sum_{k=1}^{t} \alpha_k \left\| \sum_{l=1}^{d} \mathcal{D}(\hat{\mathbf{x}}_k^l) \hat{\mathbf{f}}^l - \hat{\mathbf{y}}_k \right\|^2 + \sum_{l=1}^{d} \left\| \frac{\mathcal{C}(\hat{\mathbf{w}})}{MN} \hat{\mathbf{f}}^l \right\|^2 .$$

$$\hat{\mathbf{x}}_k^l, \hat{\mathbf{f}}^l, \hat{\mathbf{y}} \in \mathbb{C}^{MN}$$

$$\downarrow$$

$$\tilde{\mathbf{x}} = B\hat{\mathbf{x}} \in \mathbb{R}^{MN}$$

Our Approach: SRDCF

$$\check{\varepsilon}_t(\hat{f}) = \sum_{k=1}^{t} \alpha_k \left\| \sum_{l=1}^{d} \hat{x}_k^l \cdot \hat{f}^l - \hat{y}_k \right\|^2 + \sum_{l=1}^{d} \left\| \frac{\hat{w}}{MN} * \hat{f}^l \right\|^2$$

$$\hat{\mathbf{x}}_k^l, \hat{\mathbf{f}}^l, \hat{\mathbf{y}} \in \mathbb{C}^{MN}$$

$$\check{\varepsilon}_t(\hat{f}) = \sum_{k=1}^{t} \alpha_k \left\| \sum_{l=1}^{d} \mathcal{D}(\hat{\mathbf{x}}_k^l) \hat{\mathbf{f}}^l - \hat{\mathbf{y}}_k \right\|^2 + \sum_{l=1}^{d} \left\| \frac{\mathcal{C}(\hat{\mathbf{w}})}{MN} \hat{\mathbf{f}}^l \right\|^2.$$

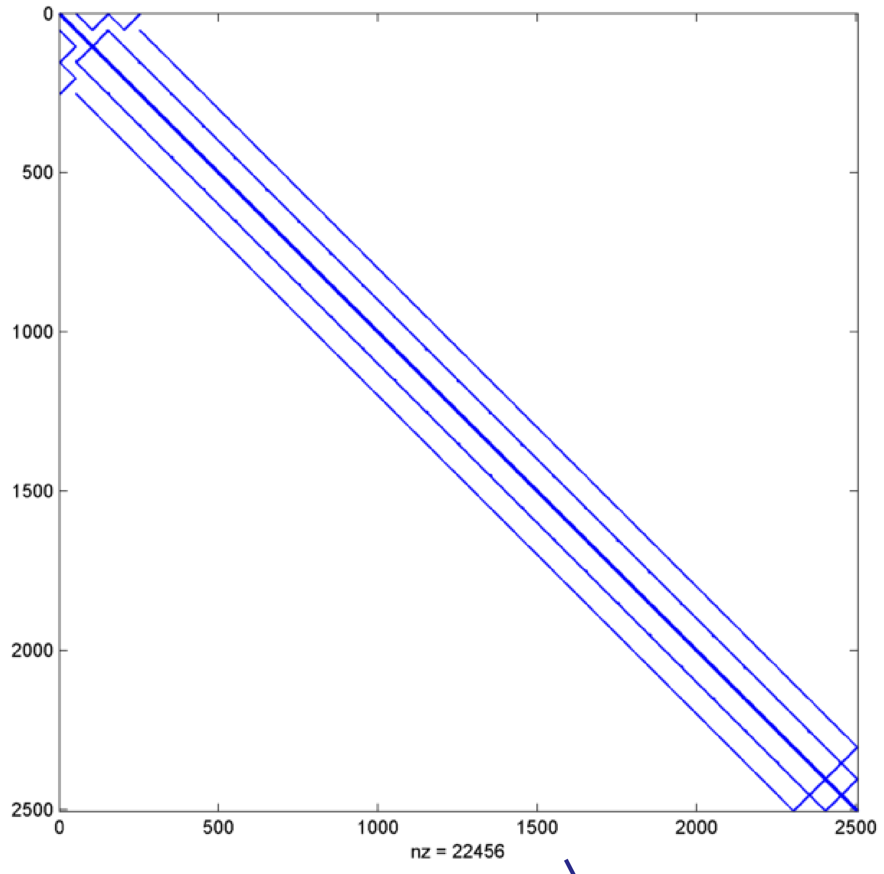
$$\tilde{\mathbf{x}} = B\hat{\mathbf{x}} \in \mathbb{R}^{MN}$$

$$\tilde{\varepsilon}(\tilde{\mathbf{f}}^1 \ldots \tilde{\mathbf{f}}^d) = \sum_{k=1}^{t} \alpha_k \left\| \sum_{l=1}^{d} D_k^l \tilde{\mathbf{f}}^l - \tilde{\mathbf{y}}_k \right\|^2 + \sum_{l=1}^{d} \left\| C \tilde{\mathbf{f}}^l \right\|^2.$$

Our Approach: SRDCF

$$\tilde{\varepsilon}(\tilde{\mathbf{f}}^1 \dots \tilde{\mathbf{f}}^d) = \sum_{k=1}^{t} \alpha_k \left\| \sum_{l=1}^{d} D_k^l \tilde{\mathbf{f}}^l - \tilde{\mathbf{y}}_k \right\|^2 + \sum_{l=1}^{d} \left\| C \tilde{\mathbf{f}}^l \right\|^2.$$

Our Approach: SRDCF



$$\tilde{\varepsilon}(\tilde{\mathbf{f}}^1 \ldots \tilde{\mathbf{f}}^d) = \sum_{k=1}^{t} \alpha_k \left\| \sum_{l=1}^{d} D_k^l \tilde{\mathbf{f}}^l - \tilde{\mathbf{y}}_k \right\|^2 + \sum_{l=1}^{d} \left\| C \tilde{\mathbf{f}}^l \right\|^2.$$

Our Approach: SRDCF

$$\tilde{\varepsilon}(\tilde{\mathbf{f}}) = \sum_{k=1}^{t} \alpha_k \left\| D_k \tilde{\mathbf{f}} - \tilde{\mathbf{y}}_k \right\|^2 + \left\| W \tilde{\mathbf{f}} \right\|^2$$

Our Approach: SRDCF

$$\tilde{\varepsilon}(\tilde{\mathbf{f}}) = \sum_{k=1}^{t} \alpha_k \left\| D_k \tilde{\mathbf{f}} - \tilde{\mathbf{y}}_k \right\|^2 + \left\| W \tilde{\mathbf{f}} \right\|^2$$

$$\downarrow$$

$$\boxed{A_t \tilde{\mathbf{f}} \; = \; \tilde{\mathbf{b}}_t}$$

Our Approach: SRDCF

$$\tilde{\varepsilon}(\tilde{\mathbf{f}}) = \sum_{k=1}^{t} \alpha_k \left\| D_k \tilde{\mathbf{f}} - \tilde{\mathbf{y}}_k \right\|^2 + \left\| W \tilde{\mathbf{f}} \right\|^2$$

$$\Downarrow$$

$$\boxed{A_t \tilde{\mathbf{f}} \; = \; \tilde{\mathbf{b}}_t}$$

$$A_t = \sum_{k=1}^{t} \alpha_k D_k^{\mathrm{T}} D_k + W^{\mathrm{T}} W$$

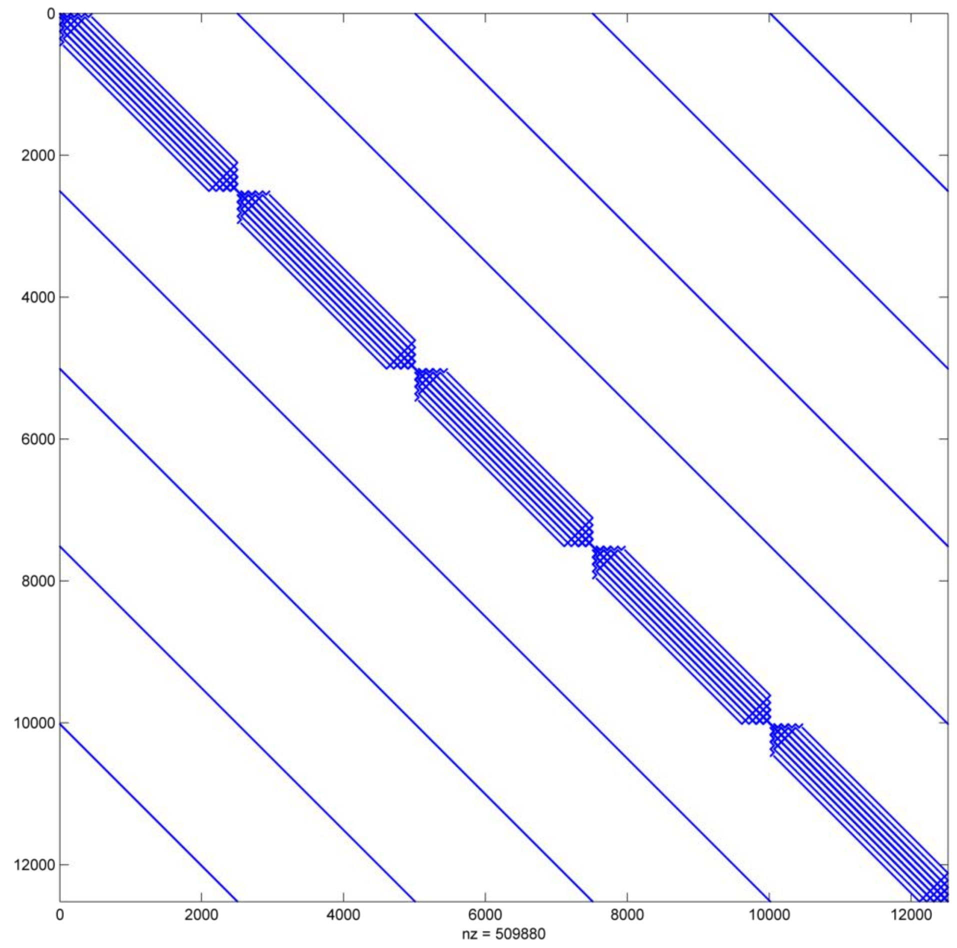
$$\tilde{\mathbf{b}}_t = \sum_{k=1}^{t} \alpha_k D_k^{\mathrm{T}} \tilde{\mathbf{y}}_k .$$

Our Approach: SRDCF

$$A_t = \sum_{k=1}^{t} \alpha_k D_k^{\mathrm{T}} D_k + W^{\mathrm{T}} W$$

$$\tilde{\mathbf{b}}_t = \sum_{k=1}^{t} \alpha_k D_k^{\mathrm{T}} \tilde{\mathbf{y}}_k .$$

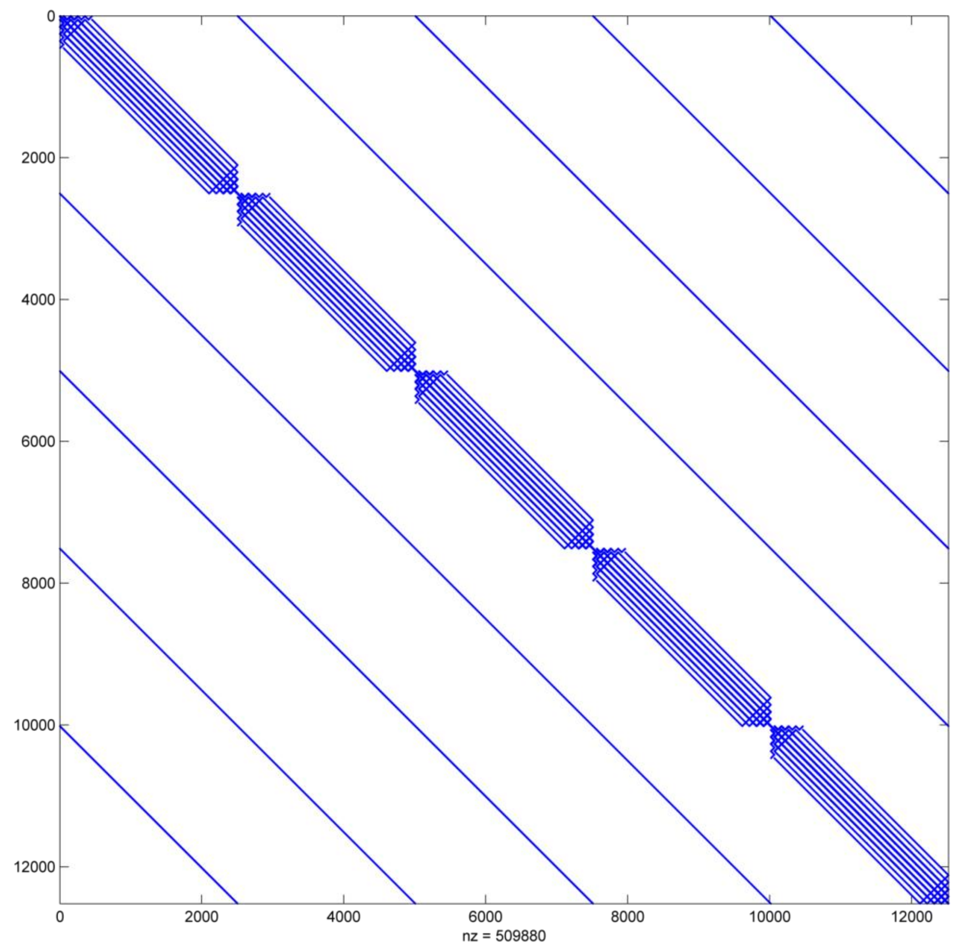
Our Approach: SRDCF



$$A_t = \sum_{k=1}^{t} \alpha_k D_k^{\mathrm{T}} D_k + W^{\mathrm{T}} W$$

$$\tilde{\mathbf{b}}_t = \sum_{k=1}^{t} \alpha_k D_k^{\mathrm{T}} \tilde{\mathbf{y}}_k .$$

Our Approach: SRDCF



$$\frac{2d+K^2}{dMN}$$

$$A_t = \sum_{k=1}^{t} \alpha_k D_k^{\mathrm{T}} D_k + W^{\mathrm{T}} W$$

$$\tilde{\mathbf{b}}_t = \sum_{k=1}^{t} \alpha_k D_k^{\mathrm{T}} \tilde{\mathbf{y}}_k .$$

Incremental Update

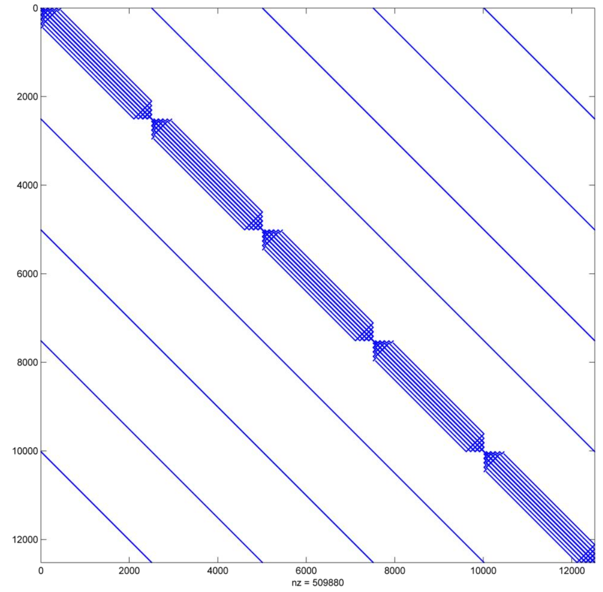
$$A_t = (1-\gamma)A_{t-1} + \gamma\left(D_t^{\mathrm{T}} D_t + W^{\mathrm{T}} W\right)$$

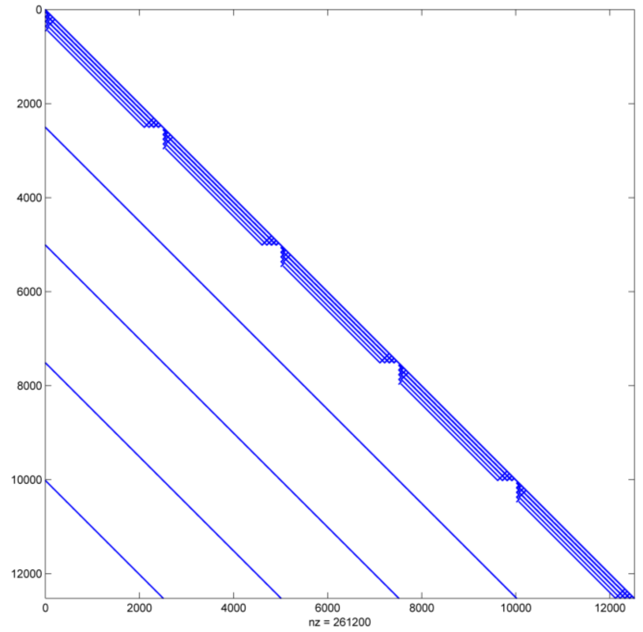
$$\tilde{\mathbf{b}}_t = (1-\gamma)\tilde{\mathbf{b}}_{t-1} + \gamma D_t^{\mathrm{T}} \tilde{\mathbf{y}}_t.$$

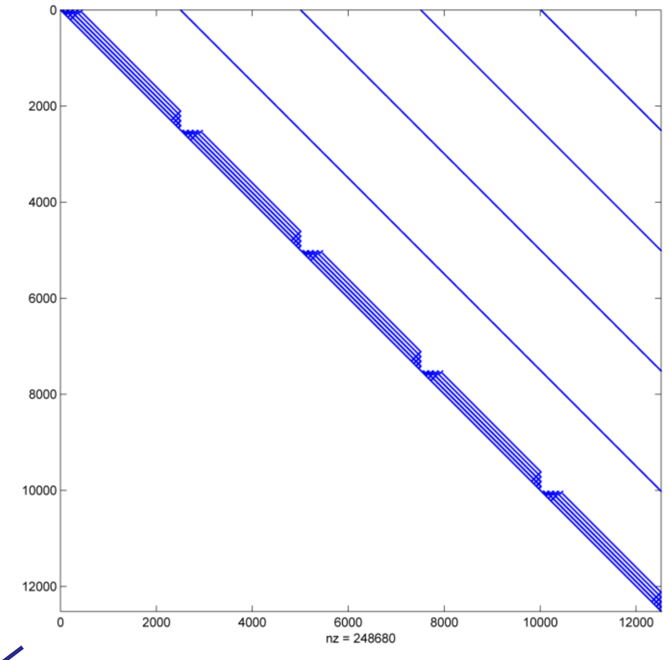
Gauss-Seidel Optimization

$$A_t = L_t + U_t$$

Gauss-Seidel Optimization

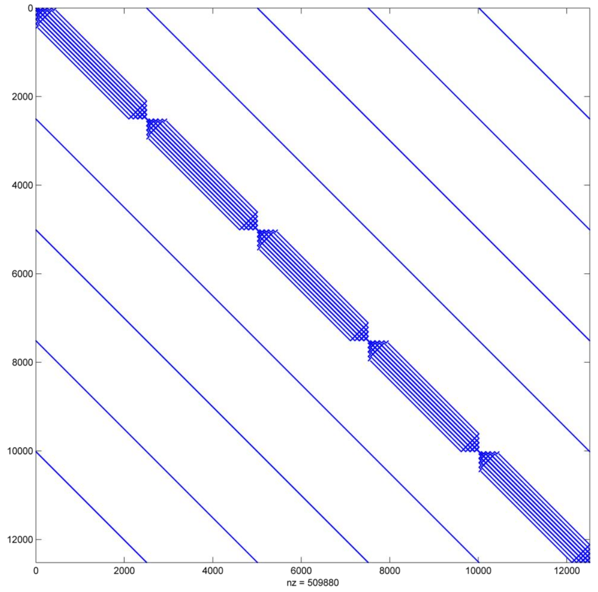


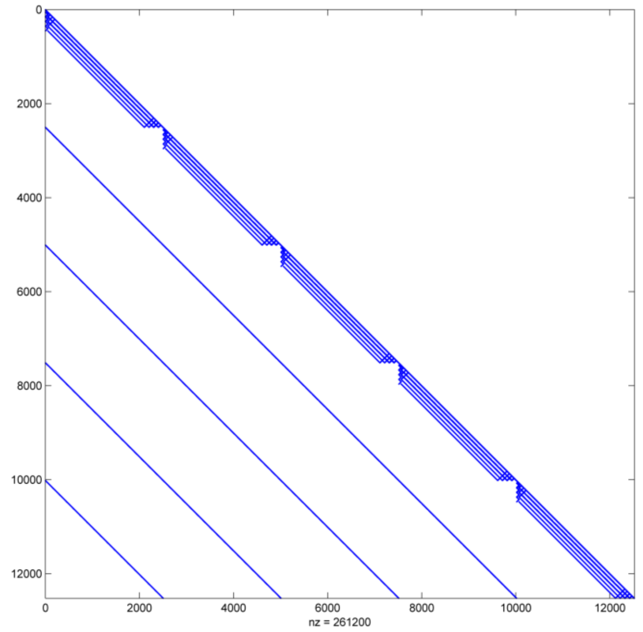


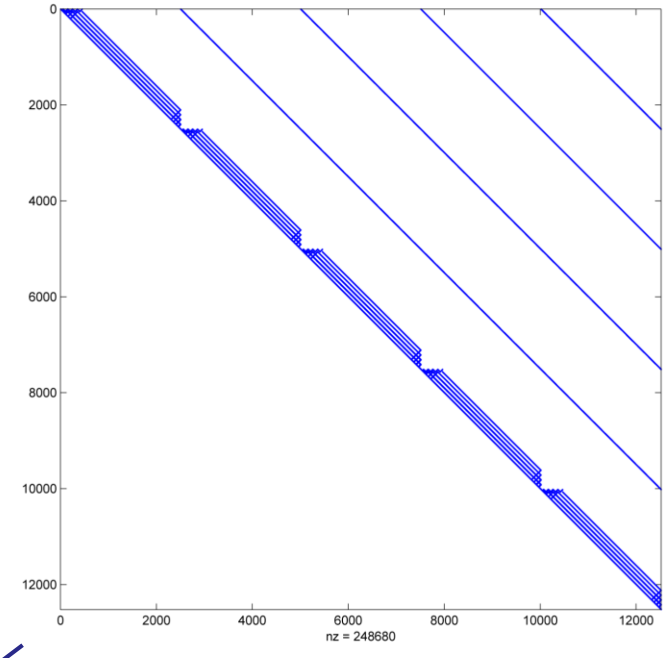


$$A_t = L_t + U_t$$

Gauss-Seidel Optimization



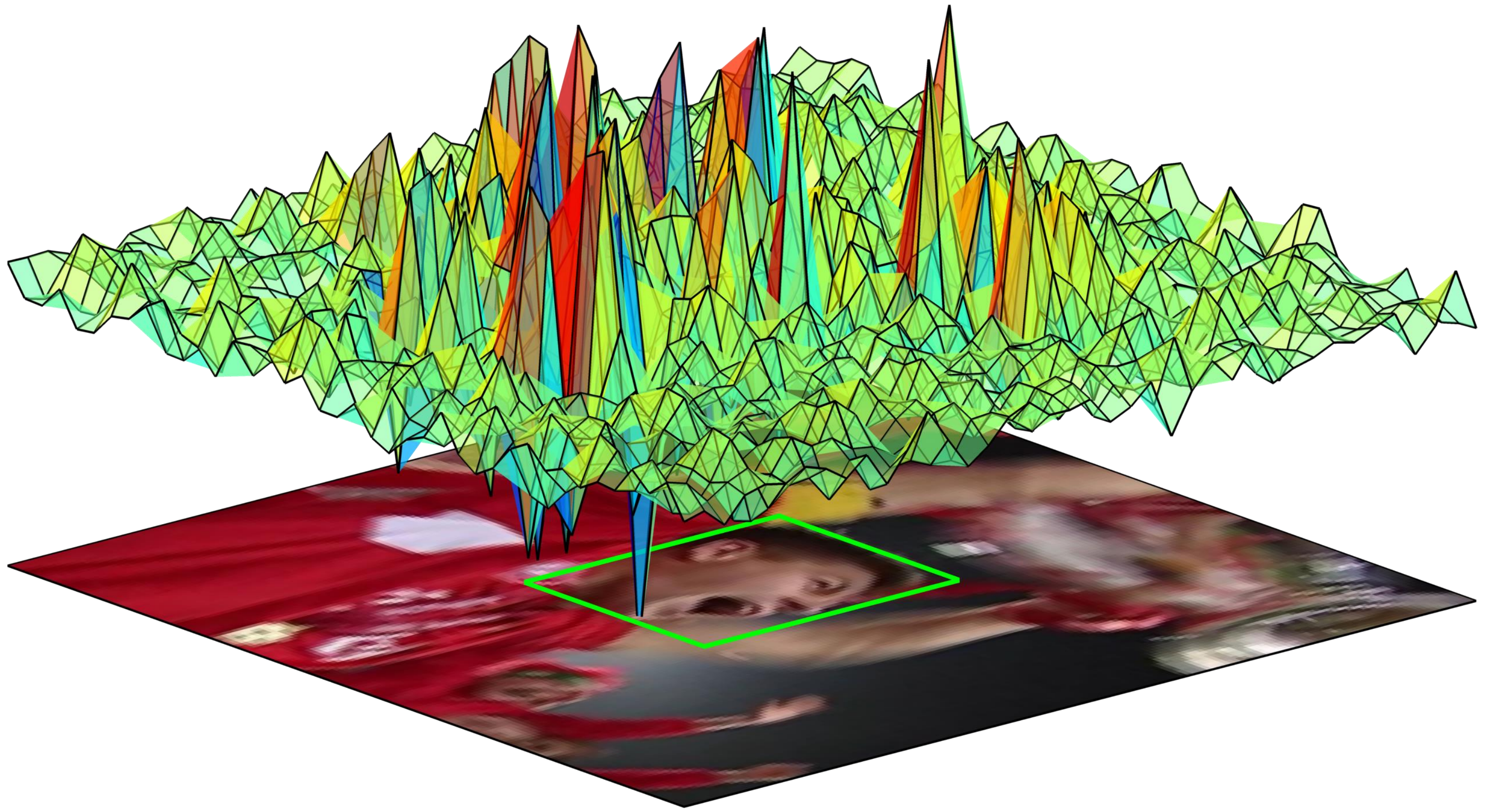




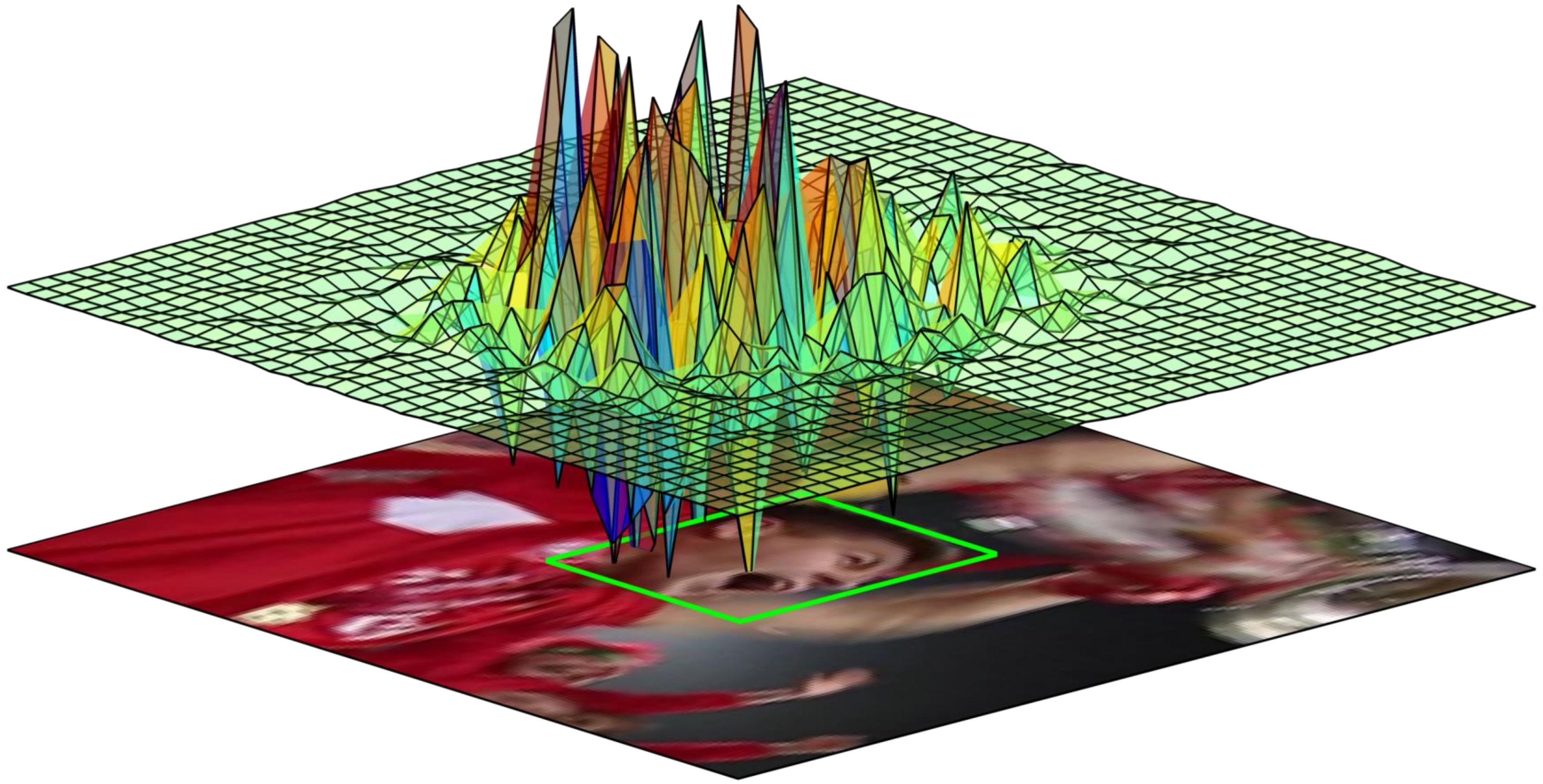
$$A_t = L_t + U_t$$

$$\boxed{L_t \tilde{\mathbf{f}}^{(j)} = \tilde{\mathbf{b}}_t - U_t \tilde{\mathbf{f}}^{(j-1)}}$$

Resulting Filter Coefficients



Standard DCF



Our SRDCF

Our Approach

Evaluation

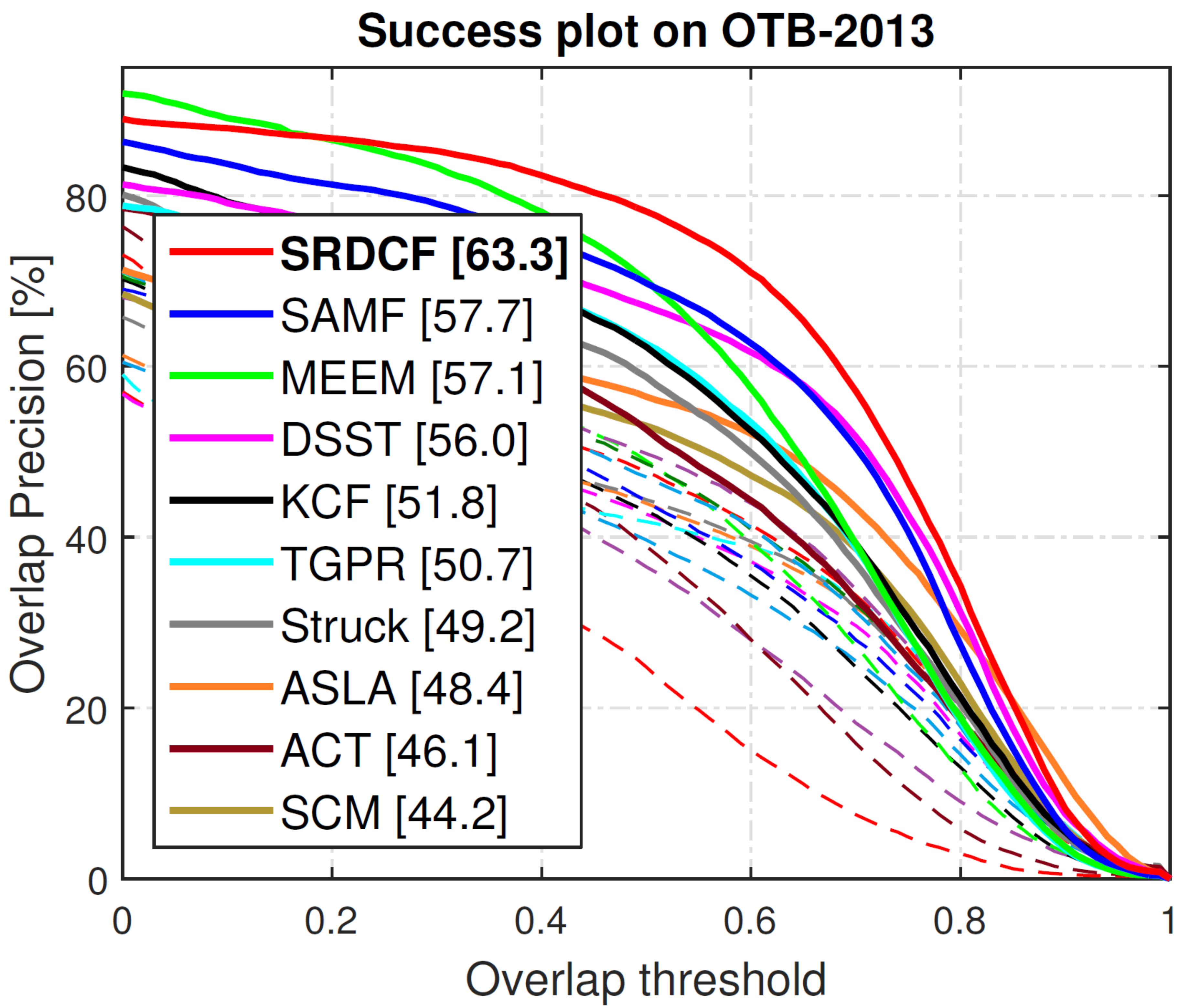
- **Four** benchmark datasets
  - **OTB-2013**, Wu et al. (CVPR 2013)
  - **OTB-2015**, Wu et al. (PAMI 2015)
  - **ALOV++**, Smeulders et al. (PAMI 2014)
  - **VOT2014**
- **VOT2015** and **VOT-TIR2015**

Impact of Regularization

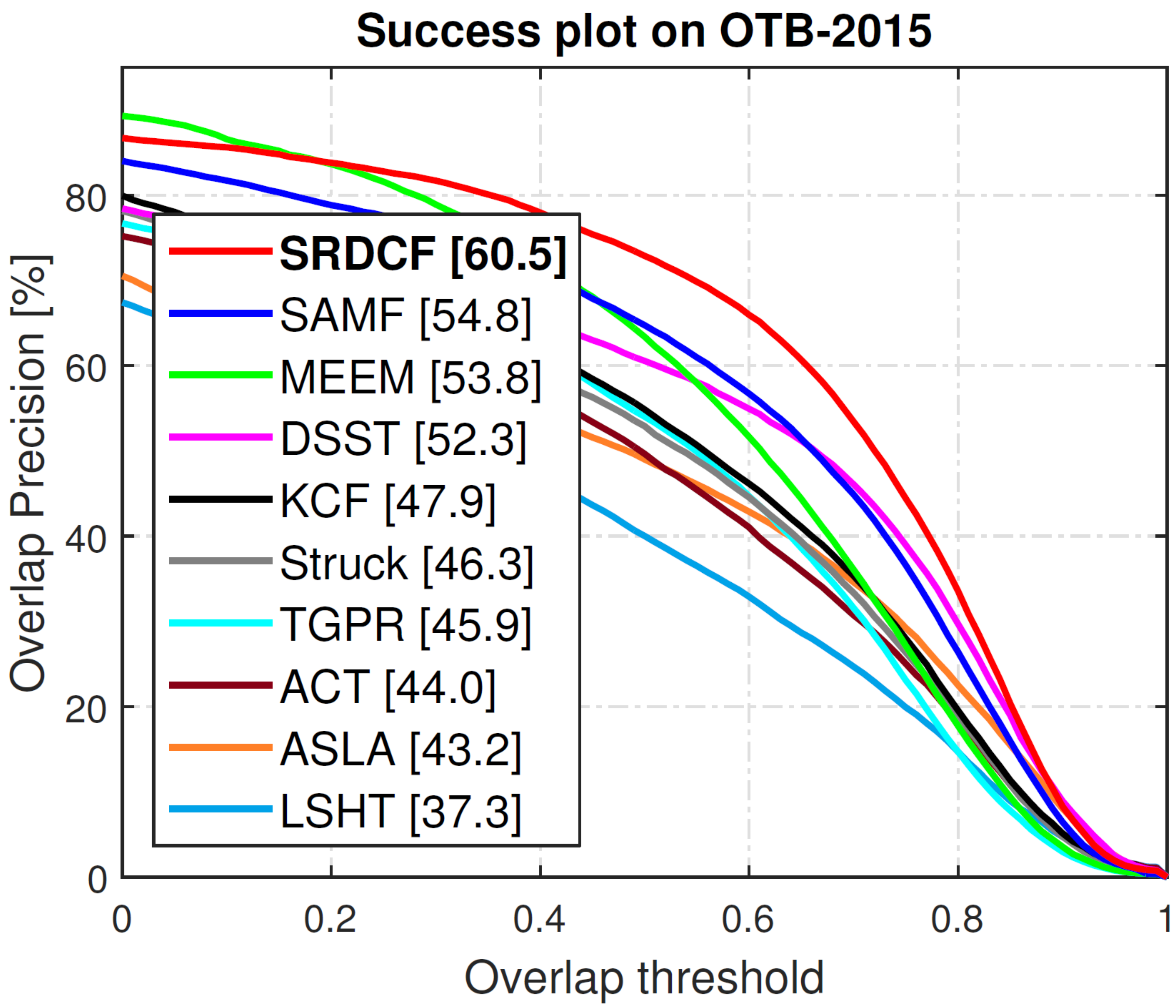
- Baseline comparison on OTB-2013
  - Our framework
  - Using HOG features

	Conventional sample size		Expanded sample size	
Regularization	Standard	Ours	Standard	Ours
Mean OP (%)	71.1	72.2	50.1	**78.1**

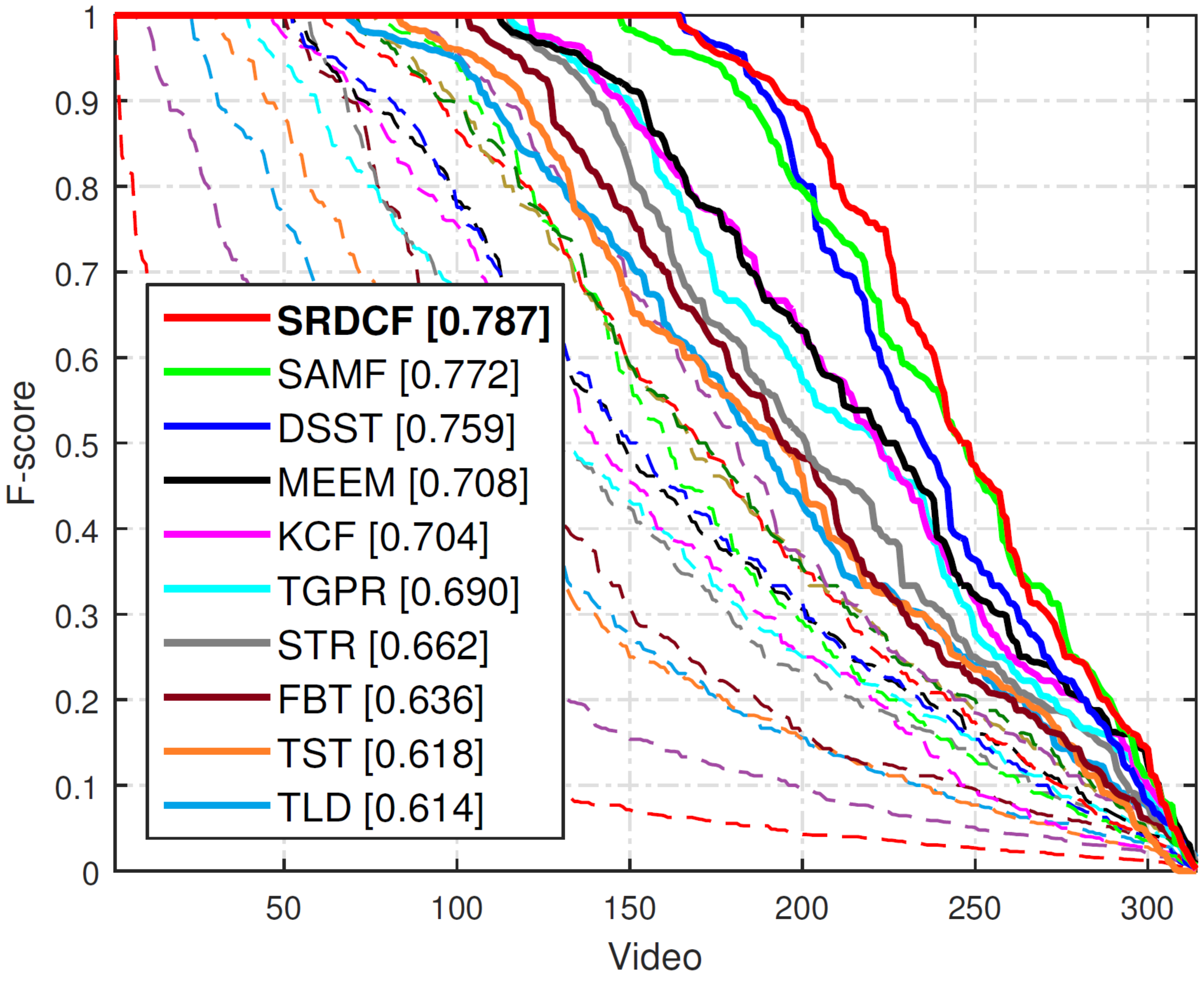
Online Tracking Benchmark 2013



Online Tracking Benchmark 2015



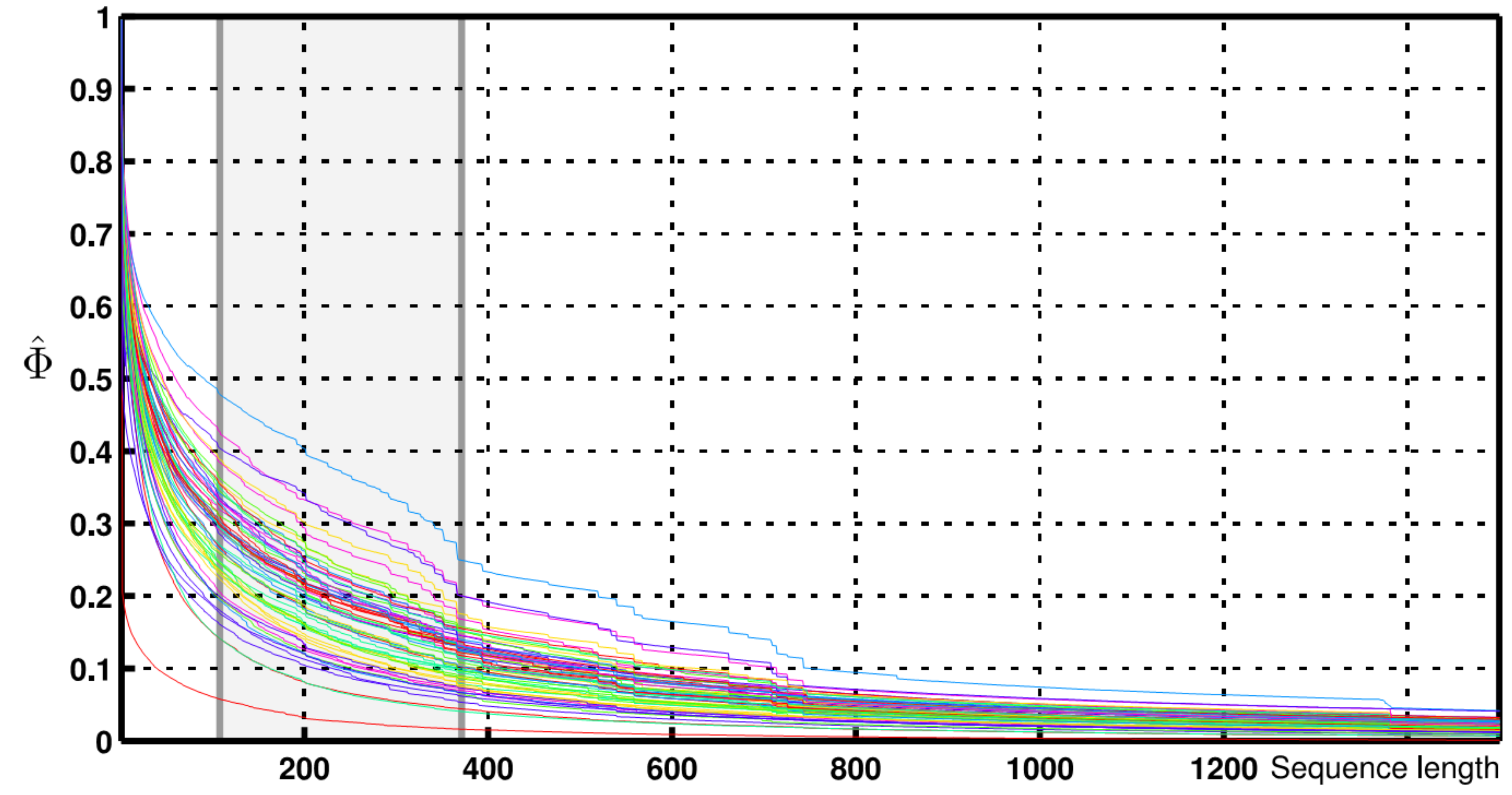
ALOV++ (314 videos)



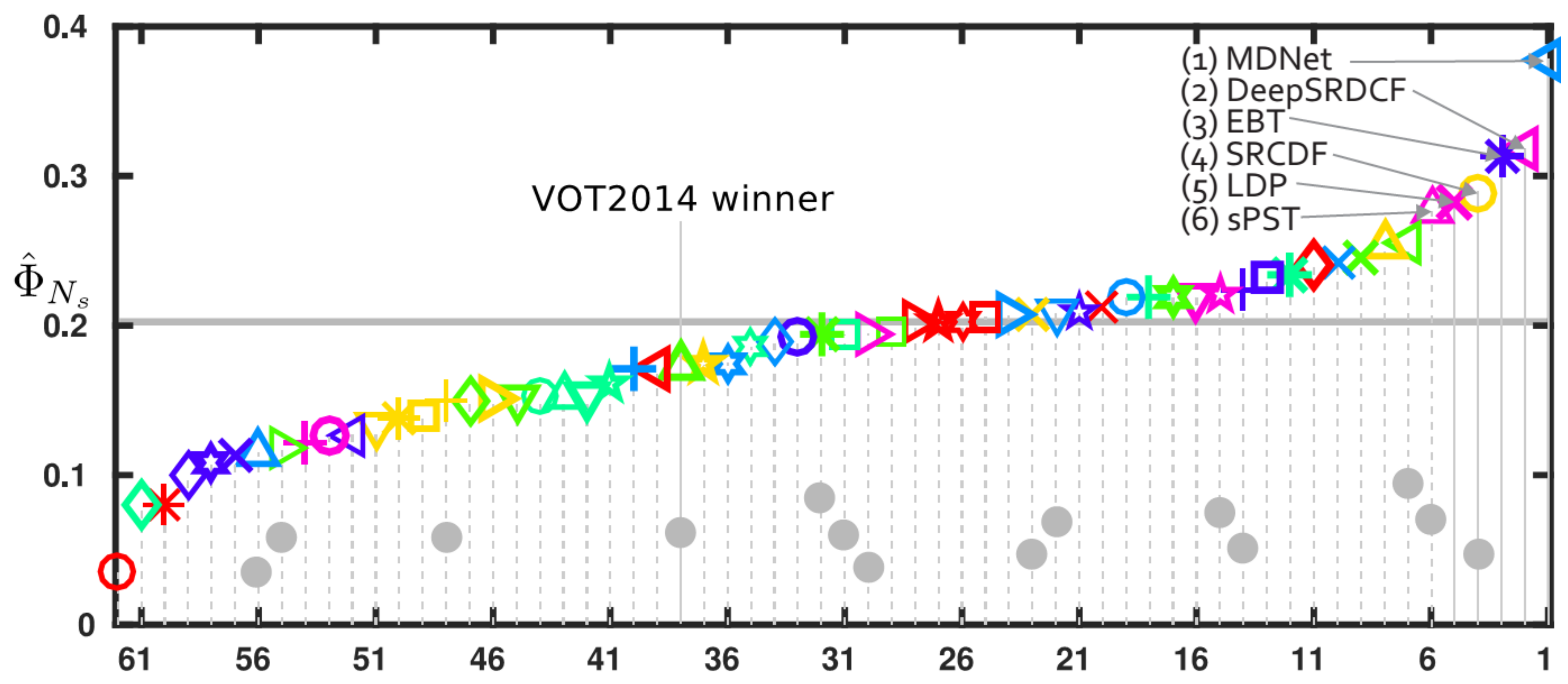
ICCV 2015 Trackers on OTB-2013

Tracker	AUC (%)
SRDCF (ours)	**63.3**
SOWP (Han-Ul Kim, Dae-Youn Lee, Jae-Young Sim, Chang-Su Kim)	61.9
Understanding and Diagnosing (Naiyan Wang, Jianping Shi, Dit-Yan Yeung, Jiaya Jia)	61.8
HCF (Chao Ma, Jia-Bin Huang, Xiaokang Yang, Ming-Hsuan Yang)	60.5
FCN (Lijun Wang, Wanli Ouyang, Xiaogang Wang, Huchuan Lu)	59.9
Proposal Selection (Yang Hua, Karteek Alahari, Cordelia Schmid)	58.0
TRIC-track (Xiaomeng Wang, Michel Valstar, Brais Martinez, Muhammad Haris Khan, Tony Pridmore)	53.0
LNL (Bo Ma, Hongwei Hu, Jianbing Shen, Yuping Zhang, Fatih Porikli)	50.8

VOT2015



$\hat{\Phi}$
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
200
400
600
800
1000
1200
Sequence length



$\hat{\Phi}_{N_s}$
0.4
0.3
0.2
0.1
0
VOT2014 winner
(1) MDNet
(2) DeepSRDCF
(3) EBT
(4) SRCDF
(5) LDP
(6) sPST
61
56
51
46
41
36
31
26
21
16
11
6
1

VOT-TIR2015

- Modified features
  - –HOG
  - –Intensity channels
  - –Thresholded frame difference