CREST: Convolutional Residual Learning for Visual Tracking

Yibing Song[1], Chao Ma[2], Lijun Gong[3], Jiawei Zhang[1], Rynson Lau[1], and Ming-Hsuan Yang[4]

[1]City University of Hong Kong [2]The University of Adelaide

[3]SenseNets Technology Ltd [4]University of California at Merced

Source code: www.cs.cityu.edu.hk/~yibisong/iccv17/index.html

International Conference on Computer Vision 2017

Introduction:

- **Single object tracking**:
  Target localization in the video frames.
- **Existing frameworks**:
  - Tracking by detection vs. discriminative correlation filter.
- **Insights on the Discriminative Correlation Filter (DCF)**:
  Pros:
  - Efficient correlation operation in the Fourier domain.
  - Dense prediction for target locations.

  Cons:
  - Boundary effect via Fourier transform.
  - The whole framework is empirically designed (i.e., filter weights training, model update, feature integration).

Our formulations:

- The objective function of DCF is ridge regression:

$$W^* = argmin_W ||W * X - Y||^2 + \lambda \cdot ||W||^2$$

- We use single convolutional layer W to replace DCF.
  - ✓ End-to-end integration with convolutional features.
  - ✓ Filter weights optimization via gradient descent.
- We adopt residual learning to measure the difference between the convolutional layer output and the ground truth.

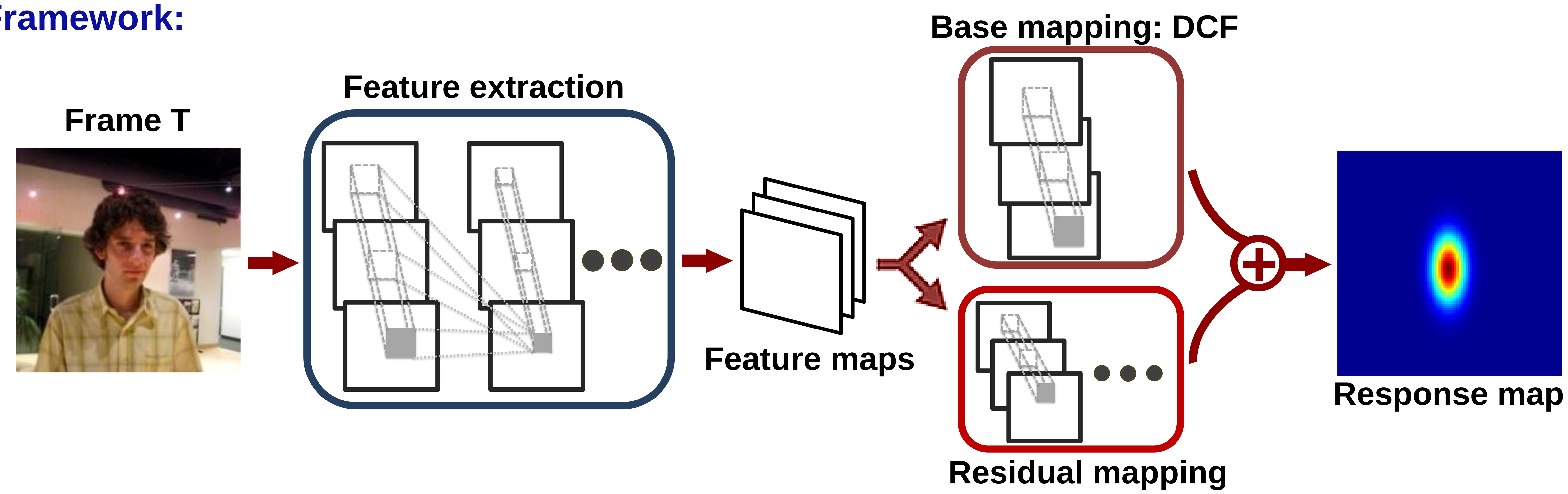
$$\mathcal{H}(x) = \mathcal{F}(x, \{W_r\}) + W * x$$

where $\mathcal{H}$ is the ground truth optimal mapping and $\mathcal{F}$ is the residual mapping.

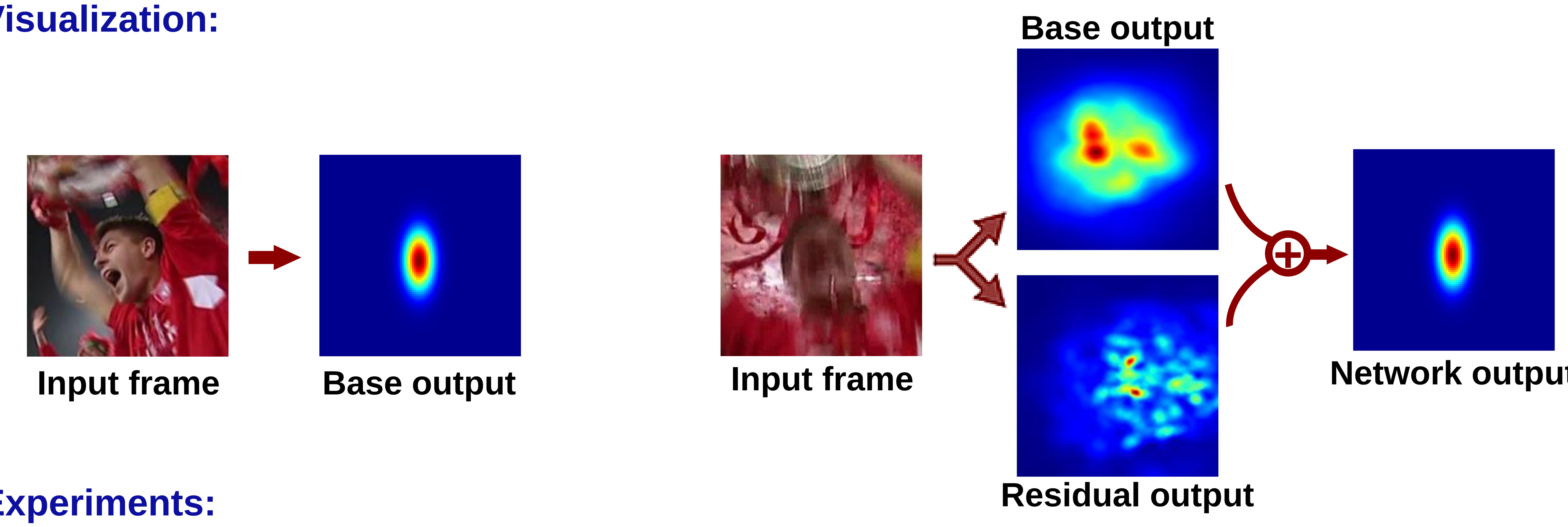
Our contributions:

- ✓ We formulate feature extraction and response generation in an end-to-end form via CNN. We adopt back propagation for model update and fully exploit the deep architecture.
- ✓ We use residual learning to handle large appearance variations, which alleviates model degradation.
- ✓ State-of-the-art performance on the prevalent benchmarks.

Framework:

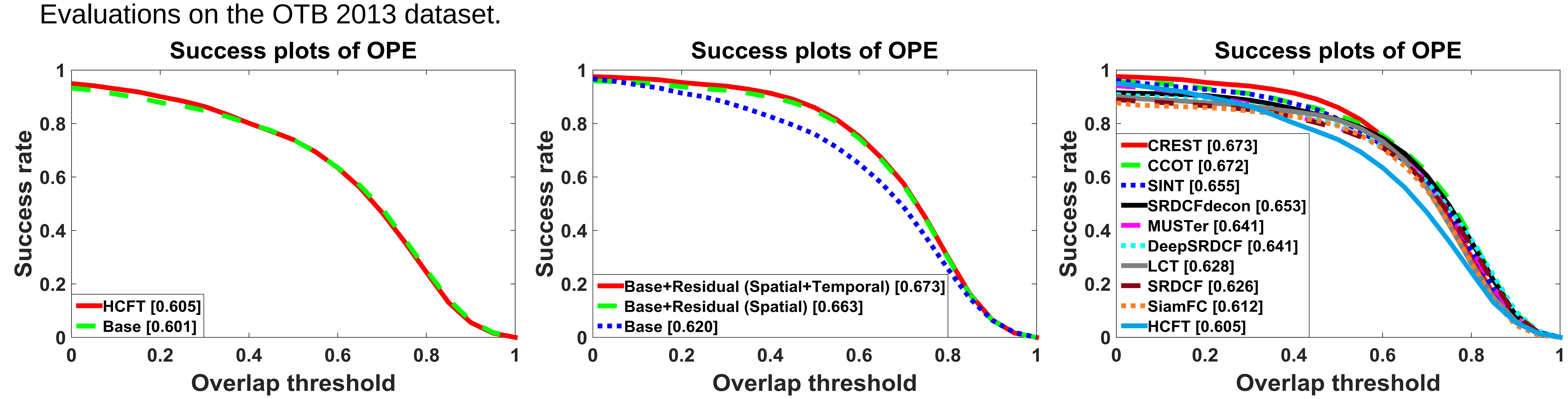


Visualization:



Experiments:

Evaluations on the OTB 2013 dataset.



Internal evaluations **External evaluations**

We show evaluations on the OTB 2015 and VOT 2016 datasets in the paper. Our implementation is available online.