Siamese FC

- 训练数据
  - 传统的跟踪方法，仅根据视频本身提供的信息来在线更新目标模板，虽然取得了一定成果，但也明显限制了跟踪模型的表达能力
- 主要思想
  - 将目标跟踪问题看作一个相似度学习问题
  - 使用大量的目标检测数据集来训练跟踪模型

Siamese FC

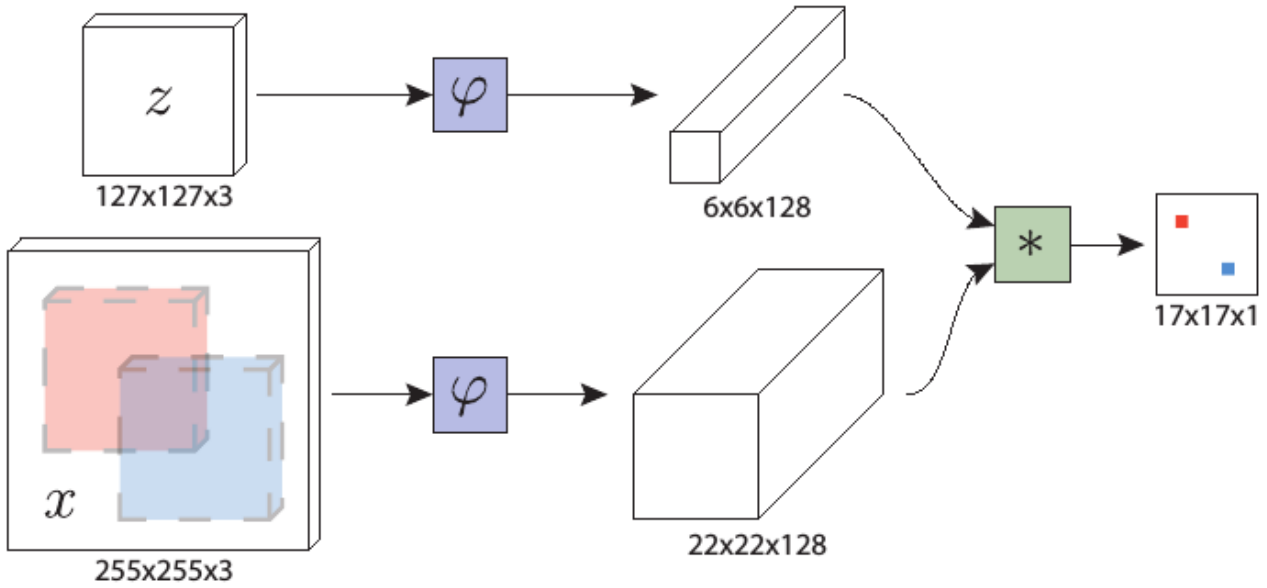

Fig. 1: Fully-convolutional Siamese architecture. Our architecture is fully-convolutional with respect to the search image x. The output is a scalar-valued score map whose dimension depends on the size of the search image. This enables the similarity function to be computed for all translated sub-windows within the search image in one evaluation. In this example, the red and blue pixels in the score map contain the similarities for the corresponding sub-windows. Best viewed in colour.

Siamese FC

- 训练细节
  - y is gt label, { +1, -1 }
  - v is real-valued score
  - L(y, v): we define the loss of a score map to be the mean of the individual loss
  - u: position in the score map
  - the elements of the score map are considered to belong to a positive example if they are within radius R of the center

$$f(z, x) = \varphi(z) * \varphi(x) + b\,\mathbb{1}$$

$$\ell(y, v) = \log(1 + \exp(-yv))$$

$$L(y, v) = \frac{1}{|\mathcal{D}|} \sum_{u \in \mathcal{D}} \ell(y[u], v[u])$$

$$\arg\min_{\theta} \mathop{\mathbb{E}}_{(z,x,y)} L(y, f(z, x; \theta))$$

$$y[u] = \begin{cases} +1 & \text{if } k\|u - c\| \leq R \\ -1 & \text{otherwise} \end{cases} .$$

Siamese FC

- 训练数据

Fig. 2: Training pairs extracted from the same video: exemplar image and corresponding search image from same video. When a sub-window extends beyond the extent of the image, the missing portions are filled with the mean RGB value.