**Motivation & Contributions**
- We present LF-Net: a novel strategy to learn a deep network for end-to-end local feature extraction pipeline (keypoints and descriptors) from raw images, from scratch.
- To do so, we propose to break the differentiability constraint present in Siamese networks, using the outputs of one network to paint a virtual target for the other.
- Ground truth (camera calibration, depth) from noisy depth sensors or off-the-shelf SfM, without human intervention.
- State of the art on wide-baseline stereo, running at 60+ fps for QVGA images. Code and models available.

**Training from whole images**
- Previous end-to-end methods (LIFT) learn from SIFT matches, which upper-bounds the performance of the keypoint detector.
- We leverage the whole image canvas and learn the optimal keypoints, along with their associated descriptors.
- The challenge: we need positive matches to train! We show how to do this by enforcing a match for each keypoint, in a non-differentiable way.

**Learning local features from patches**

**Learning local features from images**

**The Local Feature Network: LF-Net**
- A fully-convolutional detector network, which outputs scale-space score maps with an orientation for every pixel.
- Spatial Transformers to sample patches around keypoints.
- Patches are fed to the descriptor network, which outputs a 256-D feature vector for every keypoint.

**Training with two LF-Nets**
- Two copies of the network, processing two different views.
- Branch \( j \) (right) is used to generate a supervision signal for branch \( i \) (left), in a non-differentiable manner. We optimize over branch \( i \) and re-use weights for branch \( j \).
- Detector: select \( K \) points from score-map, build sharp map for \( j \), enforce \( i \) to be similar. Descriptor: warp selected keypoint locations to guarantee correspondences \( i \) to \( j \).

**Loss Functions**
- Detector: \( L_{det} = L_{im} + \lambda_{ori} L_{ori} + \lambda_{s} L_{s} + \lambda_{pair} L_{pair} \)
- Descriptor:
  \[
  L_{desc} = \sum_{k} \max(0, |D^k_i - D^k_j|^2 - |D^k_i - D^k_j|^2 + \epsilon)
  \]

**Evaluation**
- **Datasets:** Outdoors (YFCC100M) and Indoors (ScanNet).
- **Ground truth:** Camera intrinsics / extrinsics from SfM and depth from SfM or sensors. Noisy, but sufficient for training!
- **Metrics:** matching score (% of correspondences we can recover with nearest-neighbor matching).
- **Results:** SoA outdoors, close to SoA indoors.

![Diagram of LF-Net](image-url)