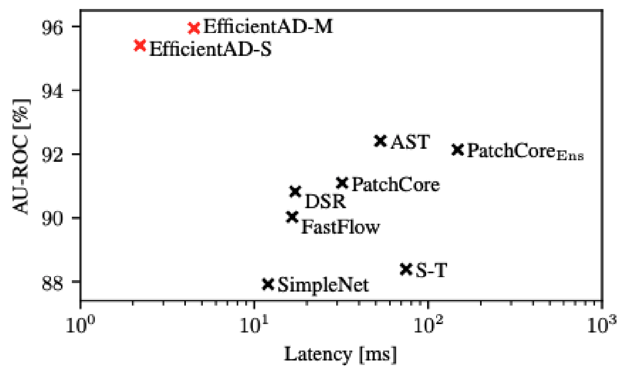
Paper Review

EfficientAD: Accurate Visual Anomaly Detection at Millisecond-Level Latencies

Why is this paper interesting?

The proposed EfficientAD method achieves high standards in both Anomaly Detection (AD) and Anomaly Localization (AD) (implementation code can be found here). It operates at a latency of two milliseconds and a throughput of 600 images per second, making it an economical and efficient solution for real-world applications. More details are shown in the sections below.



High-level summary

They have 3 main contributions they combine to achieve a proposedly new state-of-the-art on AD/AS:

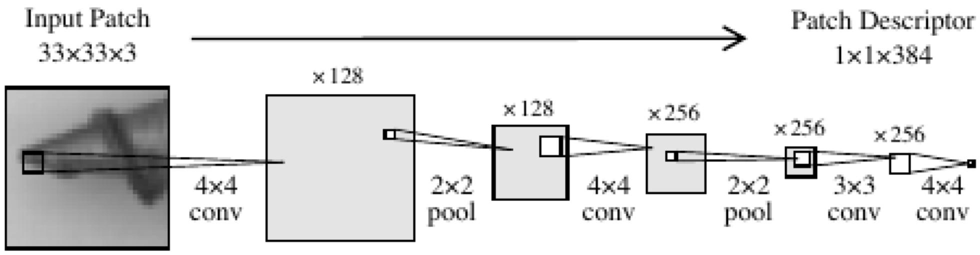
1. Efficient network architecture (Patch Descriptor Network) for fast inference/extraction of features
2. Use a specific loss-function/regularization to actually avoid the problem of "overgeneralization to anomalies" for student-teacher approaches.
   Notes: might be worth checking papers that propose a similar approach.
   a. DRÆM – A discriminatively trained reconstruction embedding for surface anomaly detection
   b. Asymmetric Student-Teacher Networks for Industrial Anomaly Detection
3. Proposal for an auto-encoder that tackles logical (i.e global) anomalies by learning the identify-function on the combination of teacher and student features.
4. Last, they propose a way to combine the auto-encoder and student-teacher scores to detect both logical/semantic and structural/textural anomalies.

Overall, the authors seem to do a lot of engineering optimisation, especially for the run-time and memory analyses.

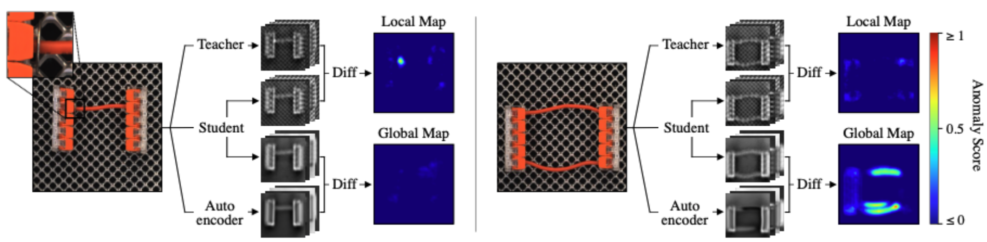
Core steps of the paper

1. Patch Descriptor Network

    a. Architecture



    b. Main benefit: we enforce short-range connections/limit the Effective Receptive Field size. => What happens if the defect is split between two patches?
2. Student-Teacher loss function has two components:
    a. Hard-negative mining loss for the knowledge distillation (i.e. only compute gradients on the X% most disagreeing feature vectors between student and teacher).
    b. Regularization term for the student on ImageNet (make the student predict all zeros to ensure that there is no generalization).
3. Auto-encoder for logical anomalies:
    a. On images with logical anomalies, the autoencoder usually fails to generate the correct latent code for reconstructing the image in the teacher's feature space.
4. Normalization of Maps
    a. Since Student Teacher loss as well as Auto-encoder Student loss are of different scales, we need a set of defect-free validation images to estimate the value ranges that can be used to normalize. This is necessary because we otherwise get False Positives
    b. Architecture



Limitations

1. They don't evaluate in the few-many shot settings, which we are mainly interested in
2. Even the method is very fast during inference, it's training step takes long time
    a. 20 mins on a strong GPU to train the whole model for each dataset.
    b. Do one-time global effort of Knowledge Distillation from a deep pre-trained classification network into the Patch Descriptor Network.
3. The method is somehow a bit complex:
    a. We don't know which hyper-parameters were hard-coded.
    b. These hyper-parameters might change depending on the different datasets.