Using Deep-Learning to Detect Video Distortions

Dr. Lev Faivishevsky
Intel Advanced Analytics

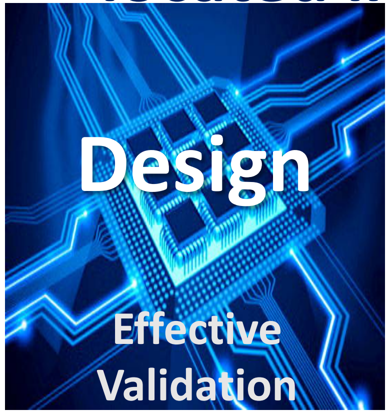
Joint work with Ronen Laperdon, Benjamin Melloul, Ravid Ziv and Amitai Armon

Intel’s Advanced Analytics department

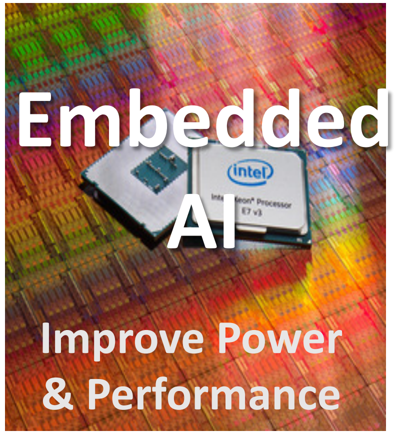
A group of 150 Data-Scientists, Big Data Developers and Product Experts + Students

located in Israel: Petach-TiKVa, Haifa, KirYat-GaT, Jerusalem

transforming Intel’s operations



Building smart products



Creating annual value of 100Ms $ using Machine-Learning / Deep-Learning / NLP / RL …

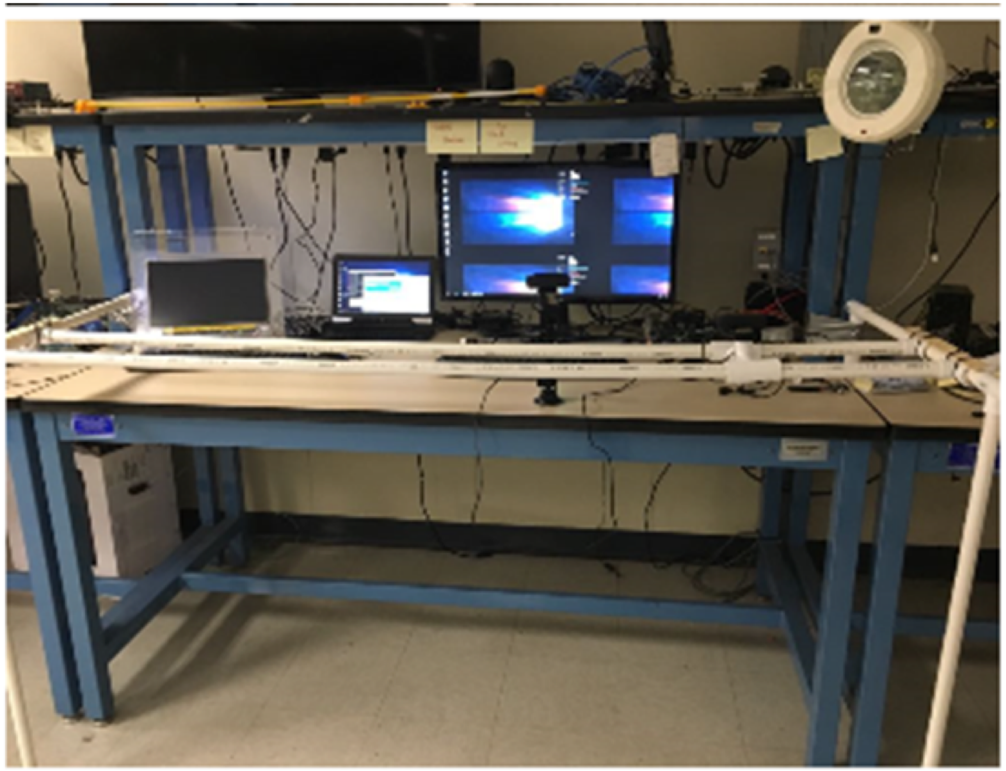
Using Deep-Learning to Detect Video Distortions

UsE-case: Validating Intel’s Processor Graphical Unit

- **Intel’s processors have a built-in graphics unit**
- **The validation of this unit and its drivers includes checking that playing movies and games looks good**
- **No-reference testing was done exclusively by people watching the screens to detect visual corruptions**

TASK: automating video corruption detection

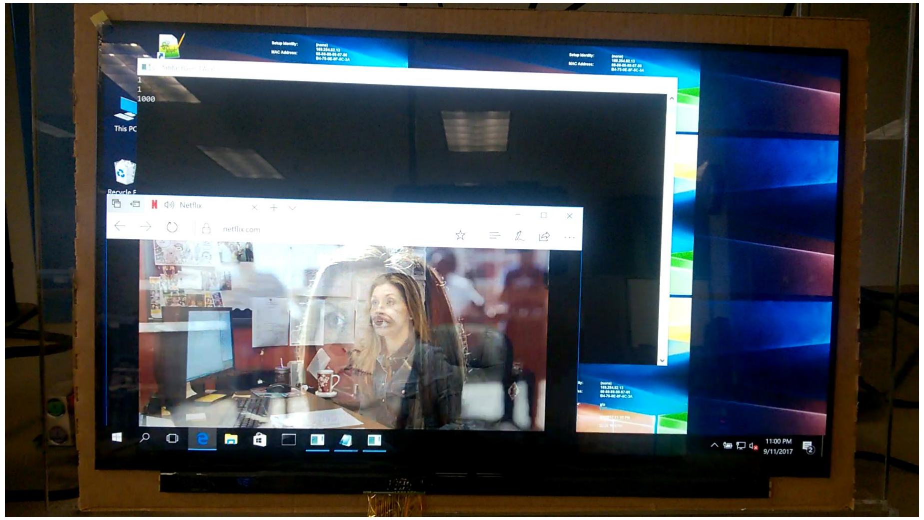
- **Goals include faster detection, increased quality and reduced costs of fixes**
- **People are intended to mainly check the system's alerts, rather than watching full movies/games**
- **Screen is captured by a camera, to enable validating the end-to-end process**



Some Challenges….

- **Need to identify noise and not objects.** Over 10 corruption types, each with several possible appearances.
- **No reference images** for the tested content (e.g., movie ads or arbitrary game actions)
- Content may include both realistic **videos and 3D games**
- Some corruptions, such as flickering, are **not seen in single frames**, only in sequences

- A 24h run on each system generates **> 2M frames**, and there are dozens, so false alert rate should be **very low**
- Avoid false alerts on lighting effects, reflections or ”ghost” effects **(see below)**
- **No collection of real data**, only few sample images of corruptions or “planted corruptions”

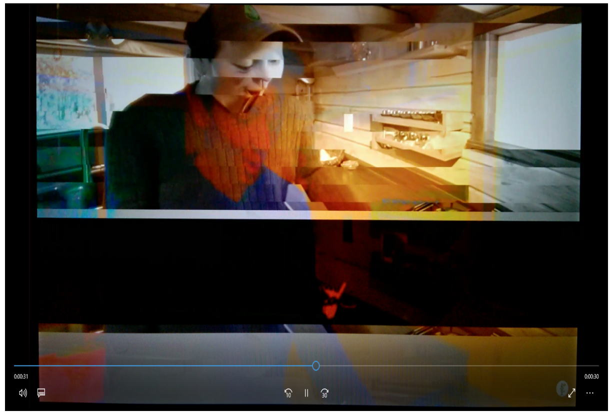


Data Collection

- **We need labeled data, but:**
  - We **cannot use** images with “planted corruptions“, nor images captured with reflections
  - New corruption cases were **quite rare** when the project started (once in few weeks)

- **Data collection approach taken:**
  - **Reproduce driver bugs** seen in the past to create corrupted data
  - For each corrupted movie, **capture an exact reference movie**, to help the model learn the noise itself
  - Additionally capture **a large amount of unlabeled good data**
  - Capture data in black boxes neutralizing reflections, with same focus

- **Data obtained:**
  - **About 5 minutes of labeled corrupted video per type**, captured with diverse content and equipment
  - Exact **reference videos** for each of the above videos + ~15 hours of **good data**



Main Modeling Approaches Tested

- **Supervised learning:**
  - Deep CNN topologies, either tuning trained models or training from scratch
  - Shallow CNN topologies on full images / patches
  - CNN with weaker supervision, considering several labels together
  - Time-based (3D) CNN on a sequence of images
  - LSTM on a sequence of features generated from images

- **Unsupervised Learning** (to cope with corruptions not included in captured data)
  - VAE based on the good data
  - Anomaly detection on image features

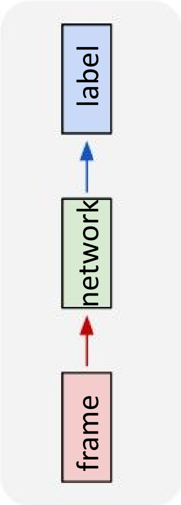
- Experiments used different data augmentations, good data sampling and image resizing

Currently the system uses an ensemble of different models:
Deep CNN (trained from scratch), Time-based CNN, LSTM on image features sequence, and anomaly detection on image features

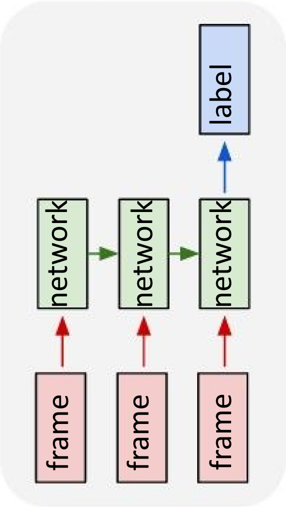
- ➢ > 80% recall on frame/sequence level
- ➢ When tuned to 0.1% false alerts

CNN DETECTION VS LSTM DETECTION

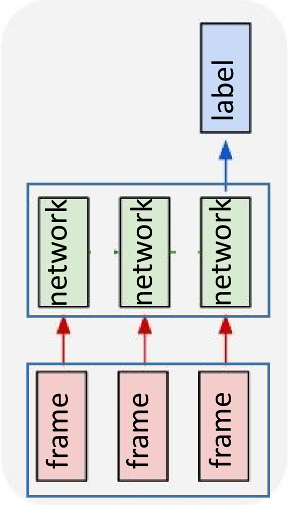
CNN



LSTM



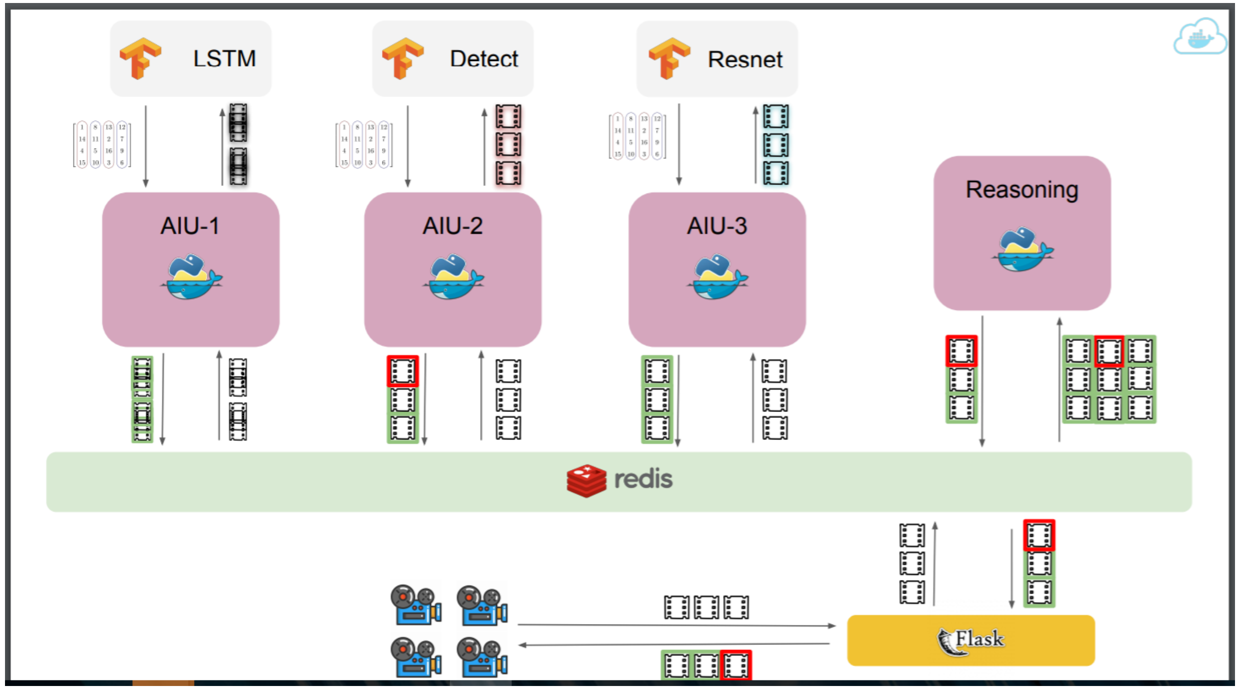
Time Based 3D CNN



- Certain corruptions (like 3D Flicker) lead to unusual sequences of frames, whereas each frame may even be not corrupted
- Need to apply model to a sequence of frames simultaneously
- LSTM models evaluate irregular sequences of frames (windows)
- Each frame is transformed into a vector of characteristics, e.g. using a network
- The time evolution of the vector of characteristics is classified to be regular or corrupted
- A score (label) is predicted for a whole sequence of frames at once

Deploying the models

- 24/7 Real-time analysis of several videos using several deep-learning models is a challenging task
- Accumulating sequences of images for analysis poses another challenge
- We built for this purpose a video analytics system that maximizes resource utilization
- **See Strata 2018 presentation by Eran Avidan: Real-Time Deep Learning on Video Streams** https://www.youtube.com/watch?v=FvJNVFKZJO0 (source of illustration below)



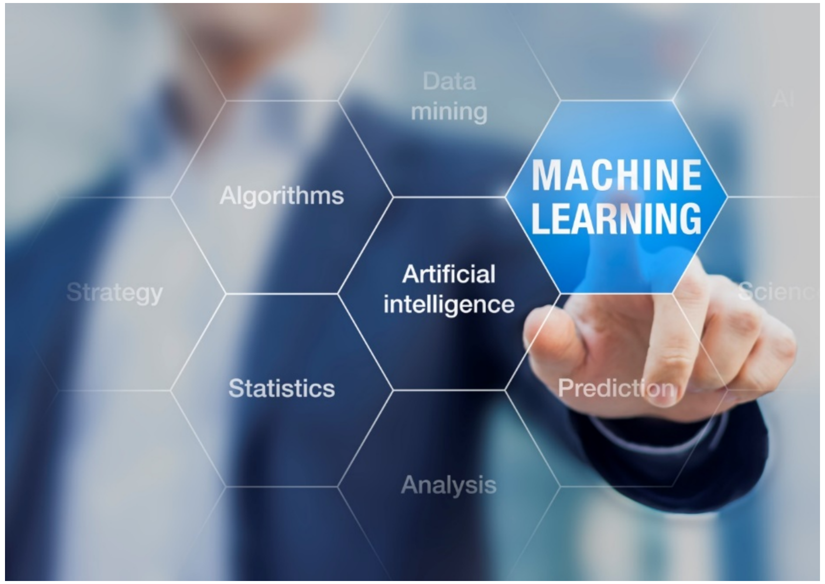
STATUS & LIVE data Results

- Side–by-side runs in progress, showing good results
- Alerts on corruption cases found so far (>10), including **one missed by humans**
- Estimated human review effort reduction is ~20x
- Deployed in two sites, full deployment planned this year

Additionally, patent application submitted

Using Deep-Learning to Detect Video Distortions

Questions?



Using Deep-Learning to Detect Video Distortions

Thank You!