Signal and Image Processing Lab.

Andrew and Erna Viterbi Faculty of Electrical Engineering
Electronics
Computers
Communications

Technion Israel Institute of Technology

Using Deep Learning to Help People with Disabilities

Yair Moshe

Deep Learning with MATLAB Seminar – May 2018, Herzliya

Who We Are

Andrew and Erna Viterbi Faculty of Electrical Engineering
Electronics
Computers
Communications

Technion Israel Institute of Technology

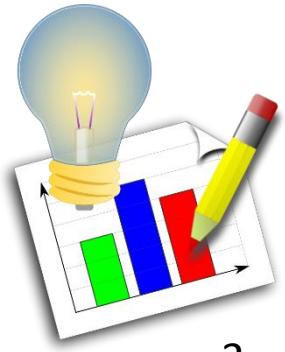
SIPL

Signal and Image Processing Lab.

- One of about a dozen labs in our department
- Established in 1975
- Fields of interest: A wide range of signal processing topics
  - Theoretical development and implementation
  - Applications for processing speech, audio, image, video, physiological signals, point clouds, and more
- Both research and education

Undergraduate Projects

- Performed by two students in their third or fourth year of studies
- Supervised by a graduate student or an experienced engineer
- Performed ‘in house’
- A taste of engineering
- Each project deals with a new challenge



- More than 50 semestrial projects each year

Undergraduate Projects

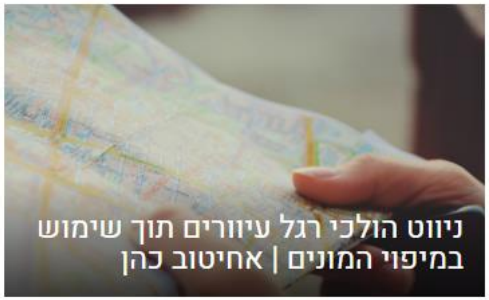
- About ⅓ of the projects are funded by industry

- Some projects end with a prototype system
- About 10% end with an academic paper
- Some end with a patent

Technion's Social Hub

החממה החברתית טכניון
social hub Technion
الدفيئة الاجتماعية التخنيون

- The flagship social engagement program of the Technion
- Strengthens the Technion's social commitment to its environment
- Initiates and supports ***SIPL*** undergraduate projects



SIPL and Deep Learning

- In recent years, deep learning has been a major tool in use of ***SIPL***
- We use various deep learning environments
  - PyTorch, TensorFlow, MATLAB

- Let's talk about two recent ***SIPL*** activities
  - Using deep learning in MATLAB
  - Help people with disabilities
  - Supported by the Technion's social hub
  - Ongoing work

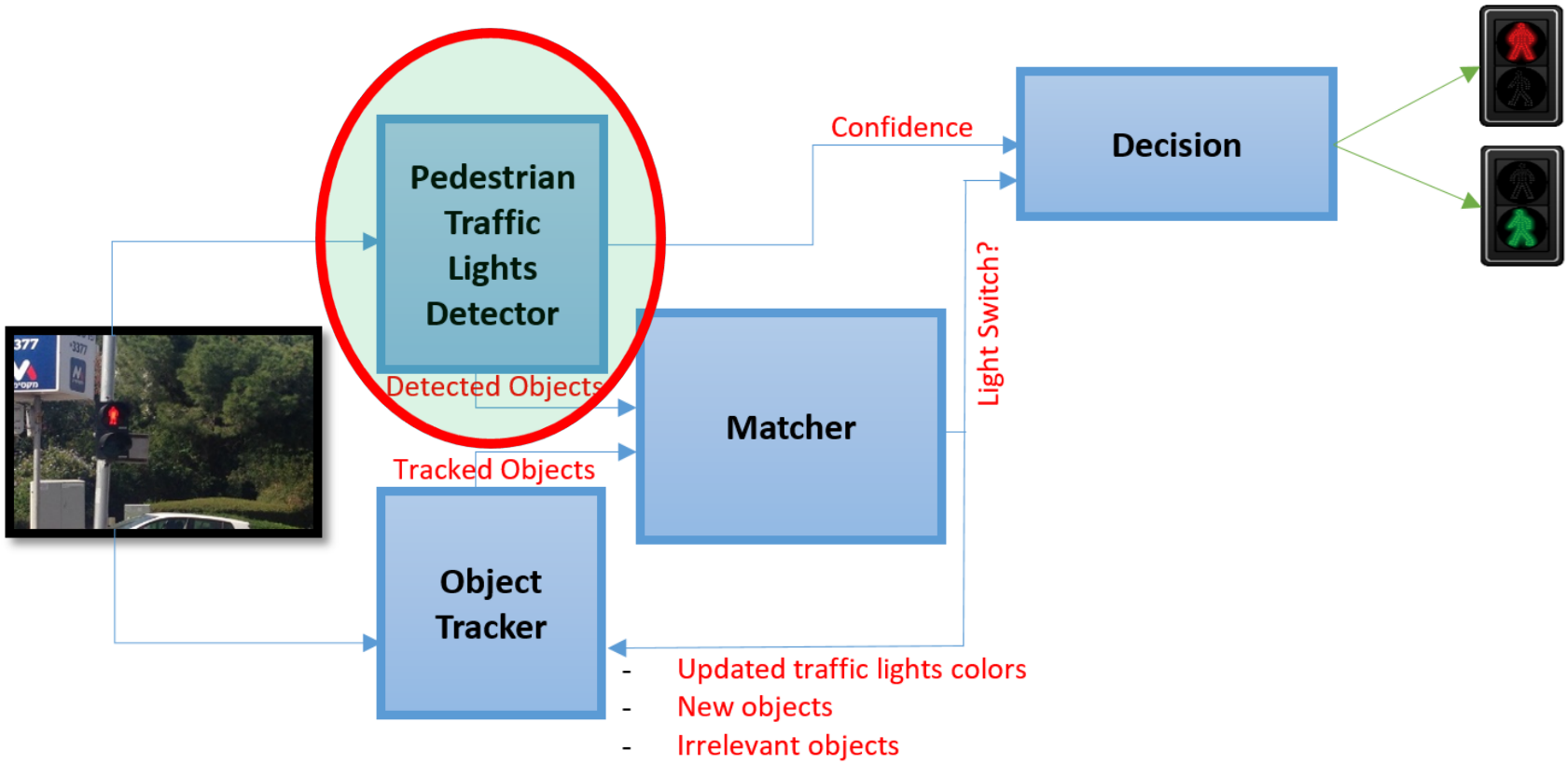
  - And we'll also talk briefly about a new lab experiment

Pedestrian Traffic Light Recognition

Idan Friedman & Jonathan Brokman
Supervised by Yair Moshe

- Many people with severe visual impairment
  - More that 250 million in the world
  - More than 24,000 in Israel
- Goal: Help these people "read" pedestrian traffic lights when no auditory cues are available
  - Develop a technique for recognizing red/green pedestrian traffic lights using video acquired from a standard smartphone
  - Implement as a mobile application

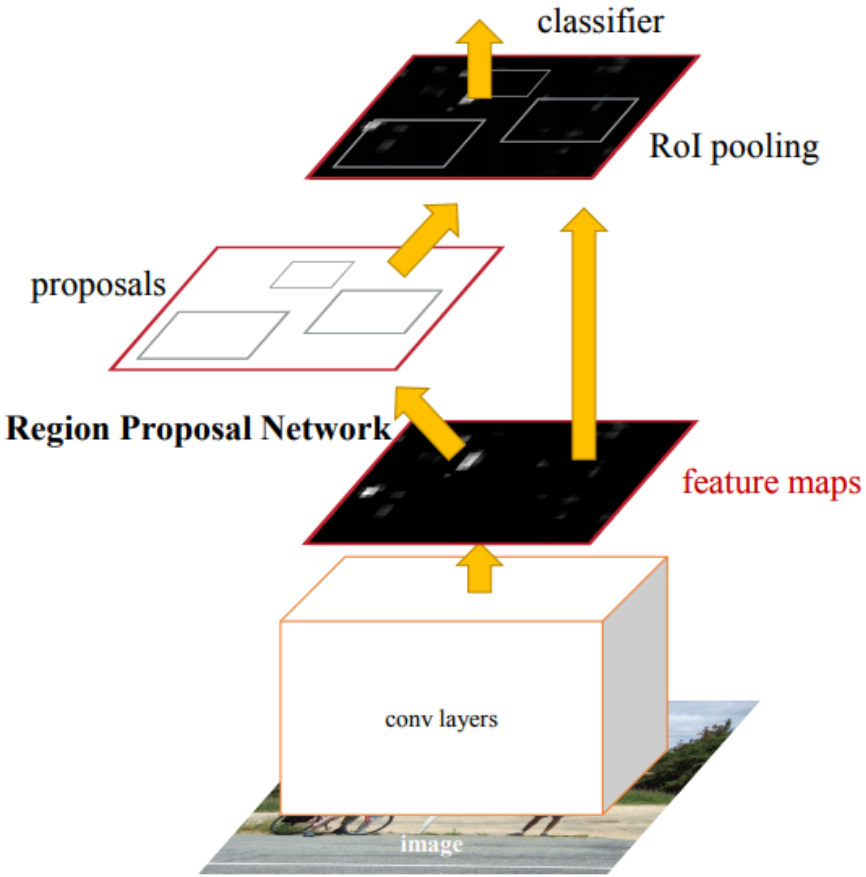
Pedestrian Traffic Light Recognition



Traffic Lights Detector

- Detection and localization per frame using Faster R-CNN - Faster Region-based Convolutional Neural Network (Ren et al., 2015)
  - Fast and accurate
  - Doesn't distort object height-width proportions
  - Trained separately for localization and classification, allowing us to give more emphasis on accurate localization
  - Built-it MATLAB implementation since R2017a

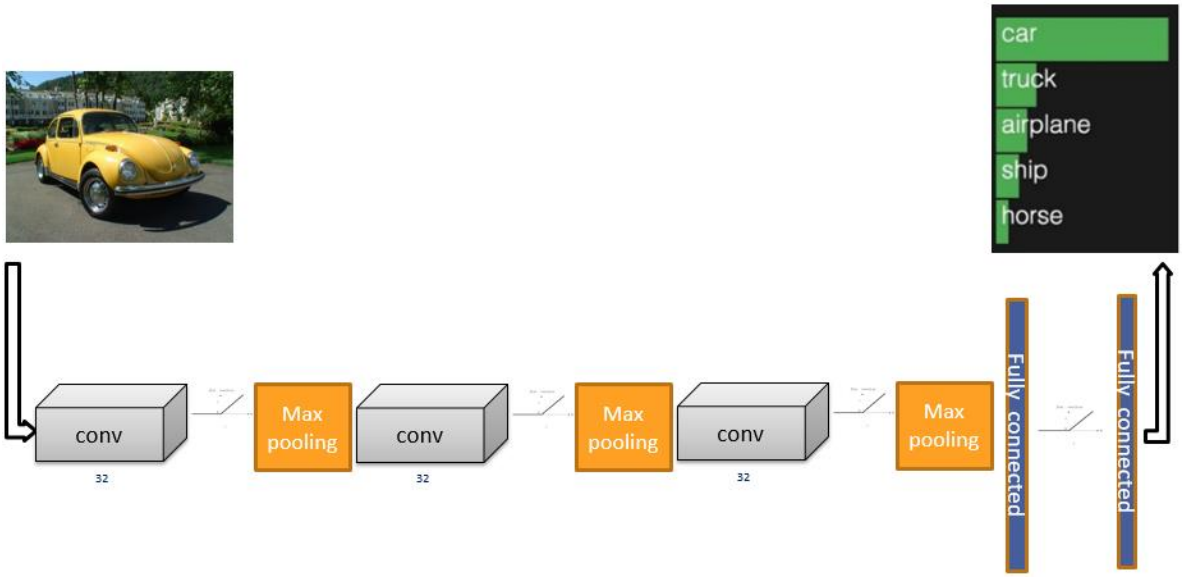
Faster R-CNN



Courtesy of (Ren et al., 2015)

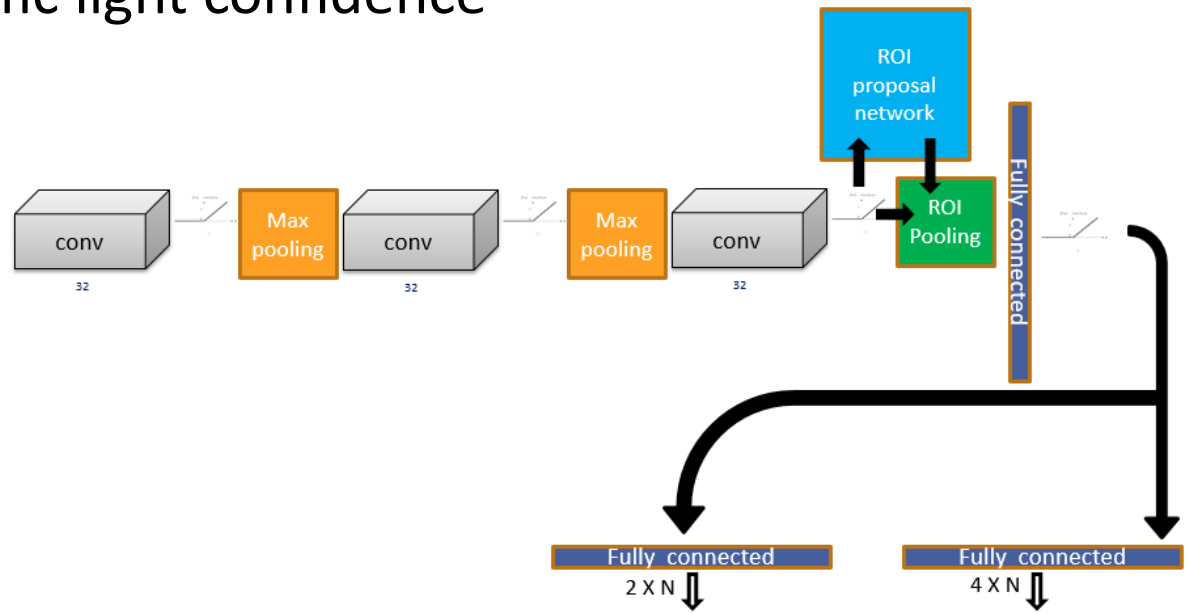
Traffic Lights Detector

- We used a network that was pre-trained for classification of 10 classes on CIFAR-10
  - [(convolution, ReLU, max pooling) x 3
    fully connected, ReLU, fully connected, softmax]



Traffic Lights Detector

- Transfer learning for pedestrian traffic lights and adaptation for Faster R-CNN
  - For each region proposal, 4 outputs for relative bounding box coordinates and 2 outputs for red/green pedestrian traffic light confidence

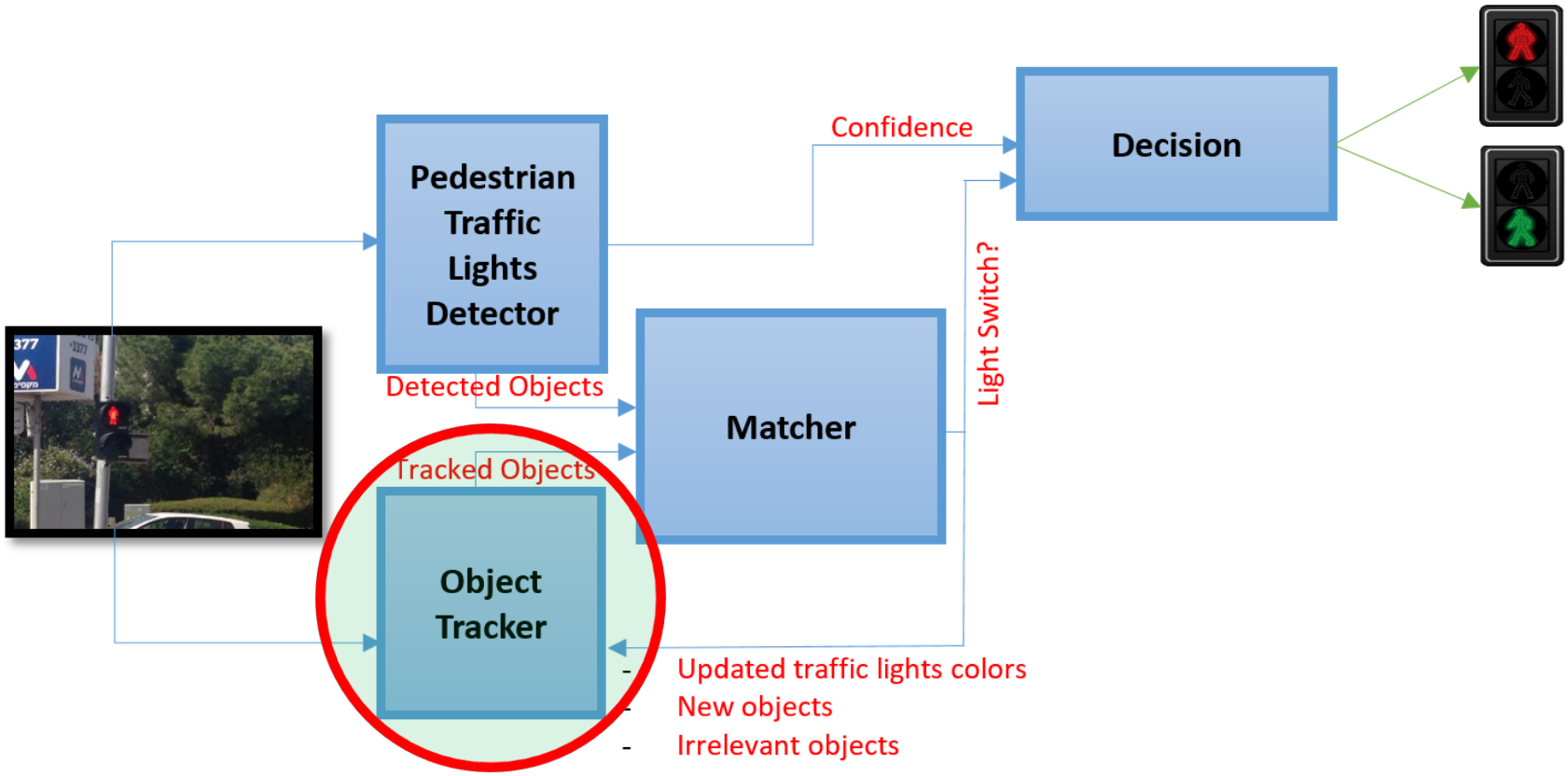


Traffic Lights Detector

- Faster R-CNN's region proposal network is trained using a combined loss function for objectness and localization

$$L_{\boldsymbol{is\ object}} + \lambda[\boldsymbol{is\ isobject}]L_{\boldsymbol{localization}}$$

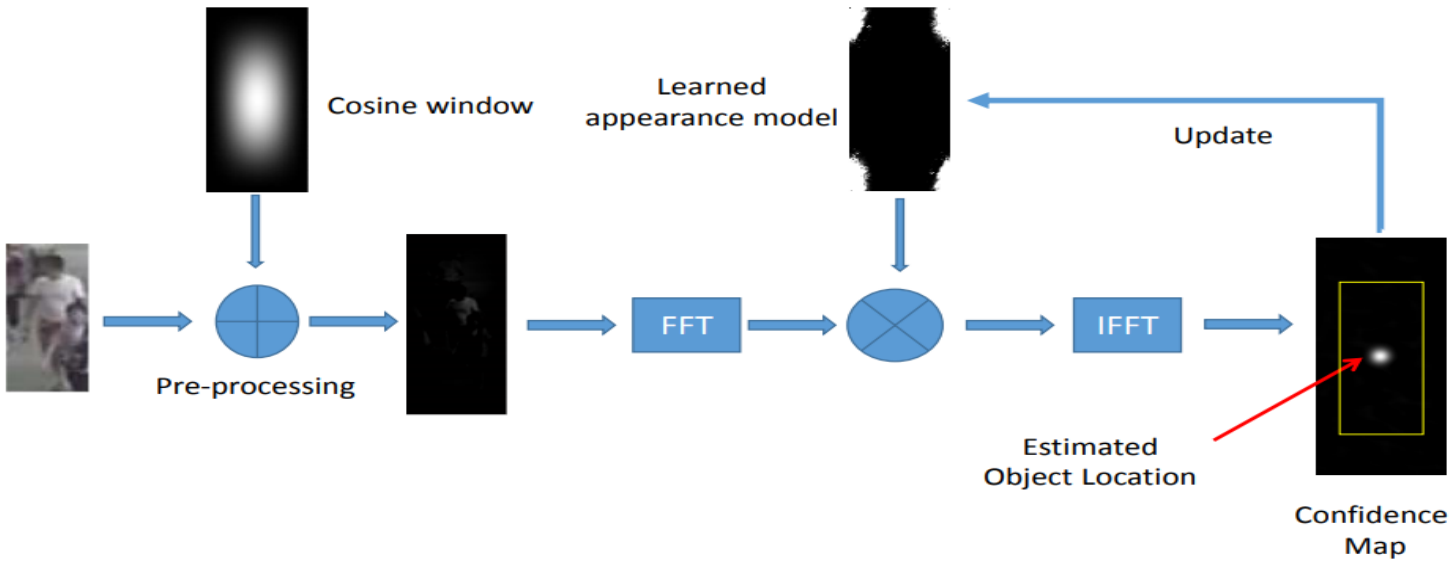
Pedestrian Traffic Light Recognition



Object Tracker

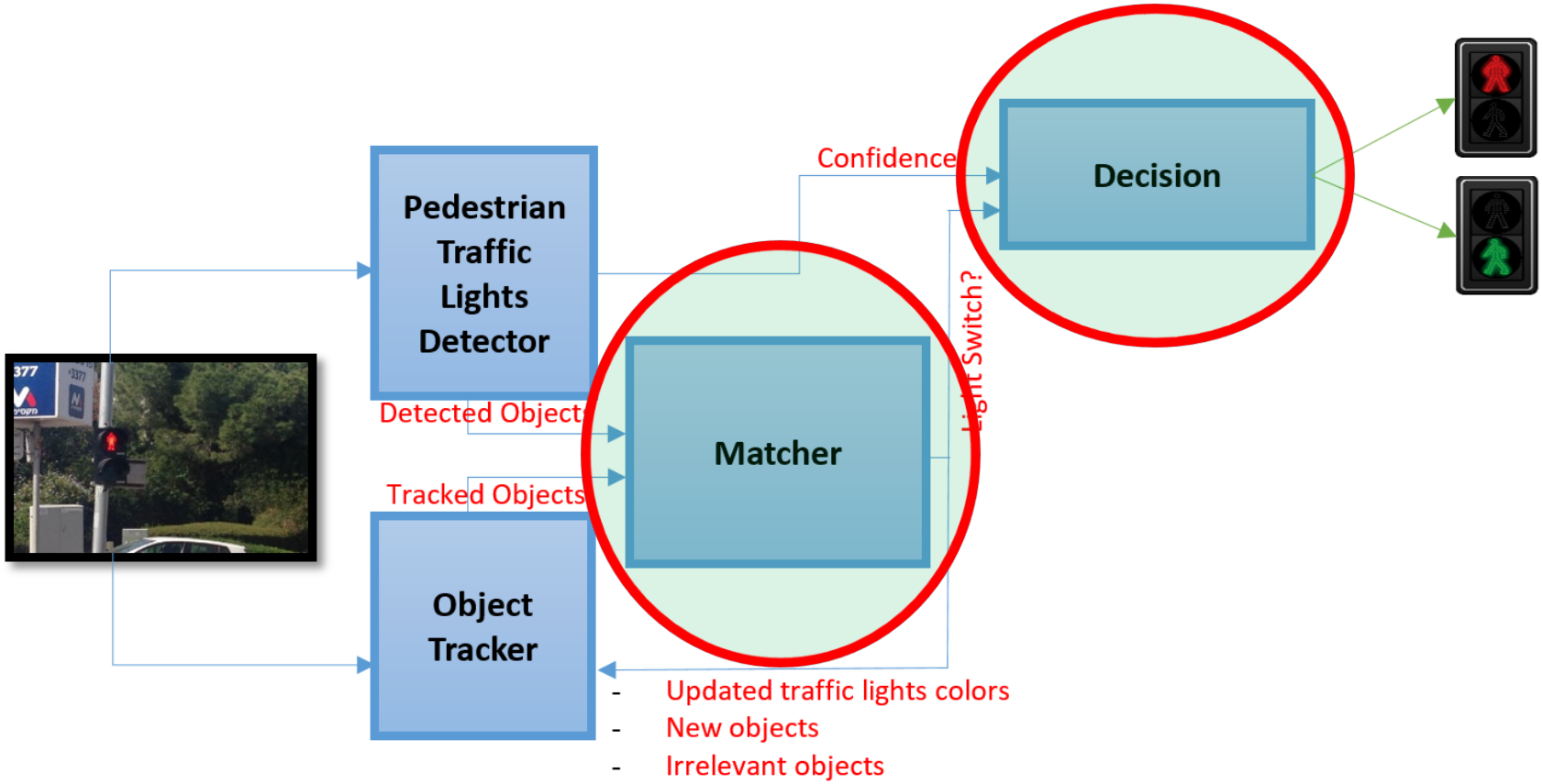
- Track region proposals along video frames using KCF - Kernelized Correlation Filters (Henriques et al., 2015)
  - Very fast and accurate
  - Author wrote a MATLAB implementation

KCF Tracker



- Based on correlation filter
  - Uses circulant matrices to efficiently incorporate multi-channel features in the Fourier domain
- Uses a Gaussian kernel and histogram-of-oriented-gradients (HOG) features

Pedestrian Traffic Light Recognition



Matching and Decision

- A detected traffic light is associated with a tracked traffic light when their bounding boxes overlap
- Final recognition occurs in one of two cases:
  1. High detection confidence score for an object that is the largest in the frame
  2. Color switch – matched tracker and detection have different red/green labels

Dataset

- Our dataset contains:
  - 950 traffic light images (450 red, 500 green)
  - 120 short videos of traffic light switches

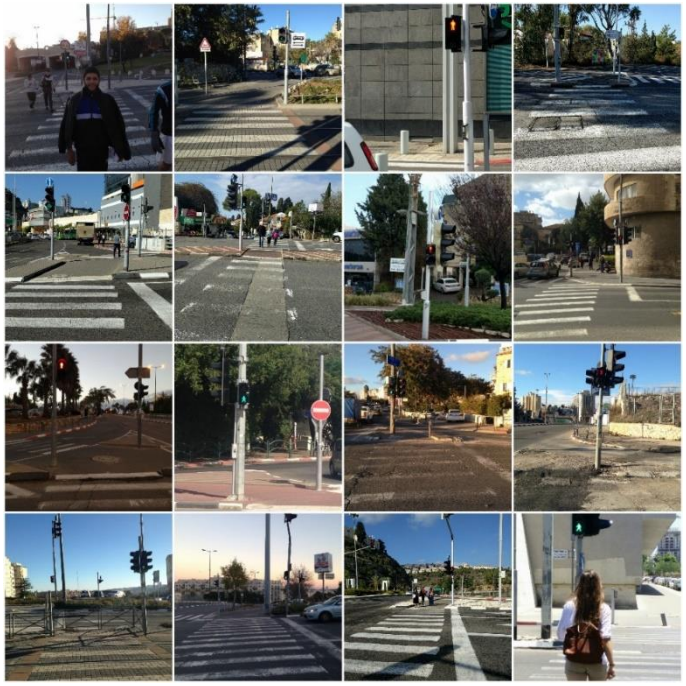


Image Recognition Results

Image Recognition Results

Image Recognition Results

Image Recognition Results

	Red	Green	Total
Recall (%)	97.6%	94.2%	95.7%
Precision (%)	98.1%	98.8%	98.5%

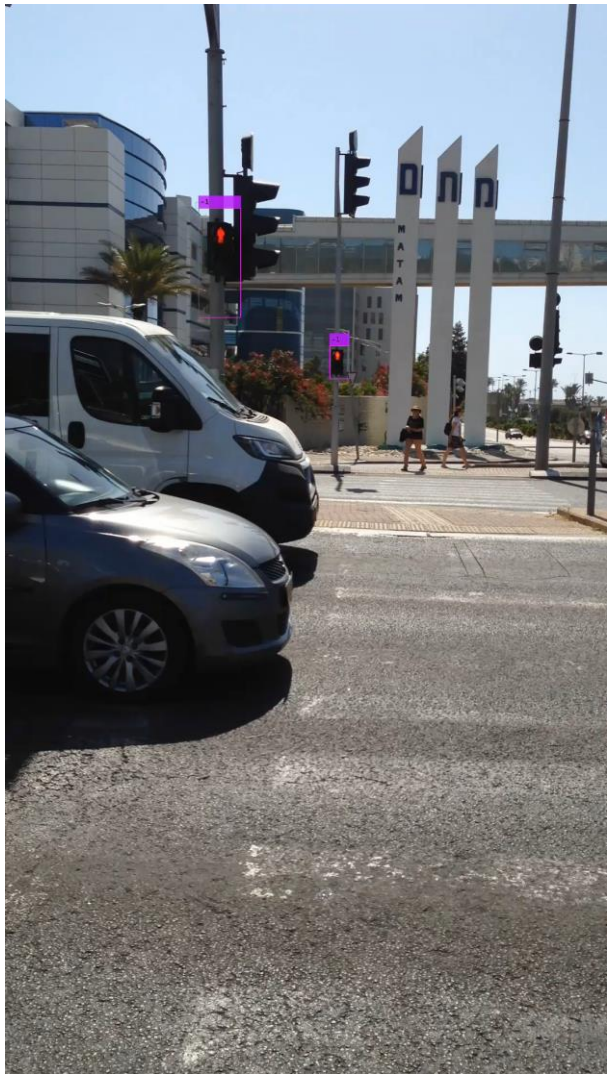
- Running time (with Nvidia GeForce GTX 1080):
  - 1600x1200: 6 frames/sec
  - 640x480: 20 frames/sec
- Nice results but not good enough for a useful application

Video Recognition Results

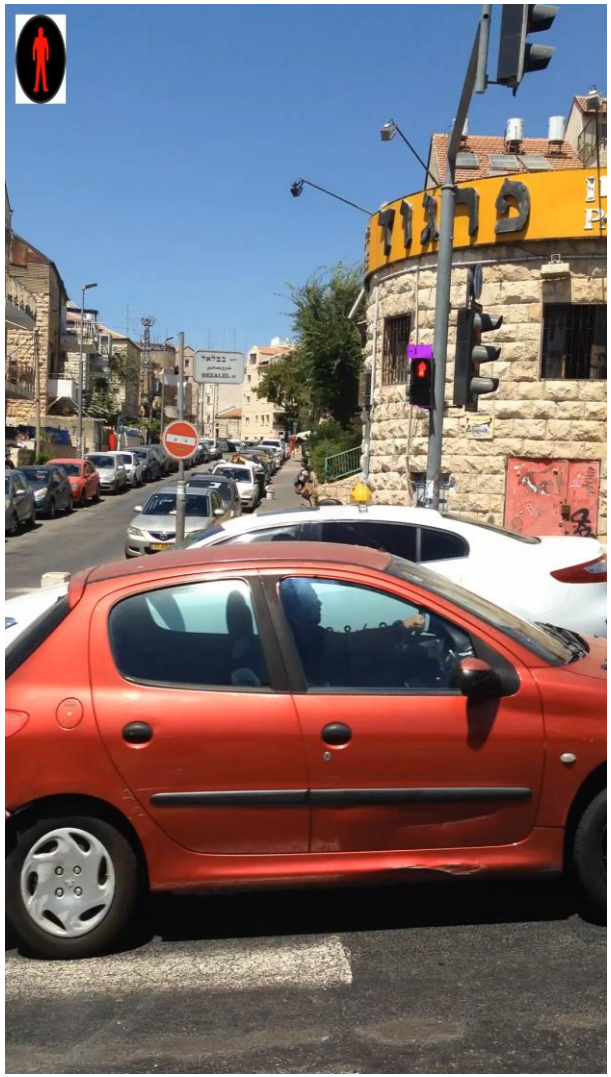
- Traffic light red/green switch detection results: 99.2%

- Switch detection is very robust and allows us to detect pedestrian traffic lights almost perfectly
- Detection is immediate or at a delay of one traffic light cycle

Video Recognition Results



Video Recognition Results



Ongoing Work

- Reimplementing as an Android application
  - Client-server architecture
- Replaced Faster R-CNN with Tiny YOLO (Redmond et al., 2016)
  - More accurate detection
  - Running at about 160 frames/sec

- Future: Incorporate scene understanding techniques

Alarm Sounds Detection

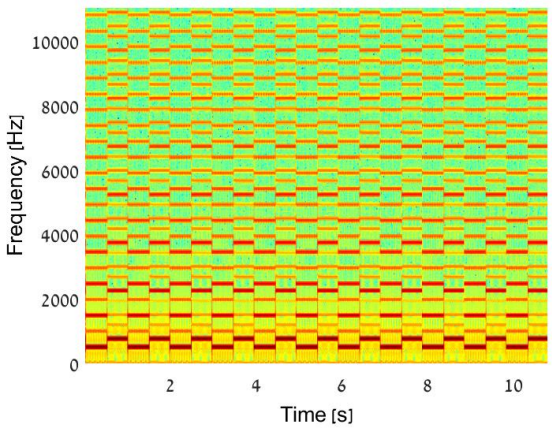
Dean Carmel & Ariel Yeshurun
Naor Fadida & Yoni Lederman
Supervised by Yair Moshe

- Many people with hearing impairment
  - About 18%
  - About 5% with moderate to severe disability
- Goal: Help these people “hear” alarm sounds
  - Develop a technique for detecting alarm sounds
  - Implement as a mobile application

Difficulties

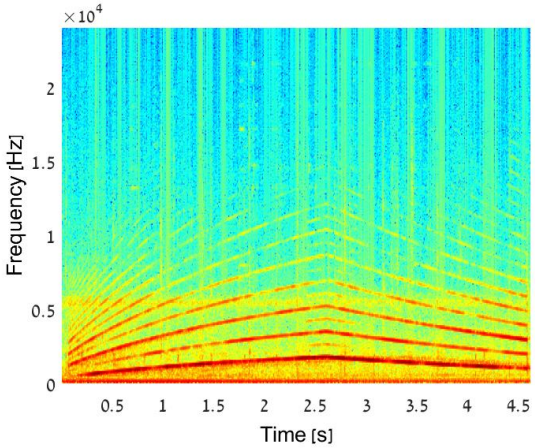
- Distinctive characteristics of alarm sounds not formally defined
  - Most countries have their unique siren standardization
  - Many alarm sounds are not standardized
    - E.g. alarm clocks
- Strong ambient noise
- Doppler effect
- Negligible false alarm rate is a must

Alarm Sounds



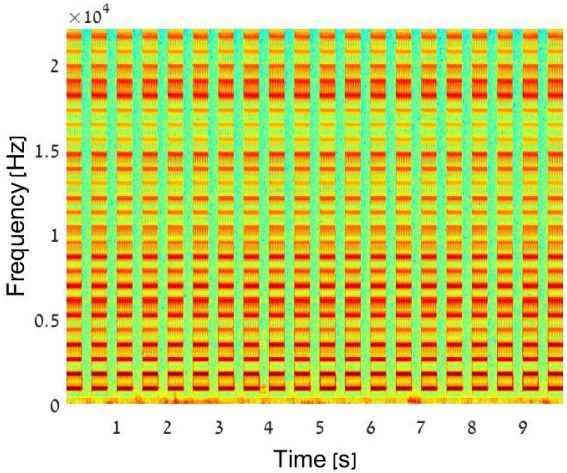
Pulsed alarm

Alarm clock



Siren

Ambulance driving away



Alternating alarm

Fire alarm

Previous Work

- Environmental audio classification
  - E.g., explosions vs. door slams vs. dog barks
- Detection of abnormal audio events
- Detection of particular alarms
  - E.g., sirens of emergency vehicles of a specific country

- We aim to design a generic alarm sound detection technique
  - Detect most electronically generated alerting sounds

Previous Work @ *SIPL*

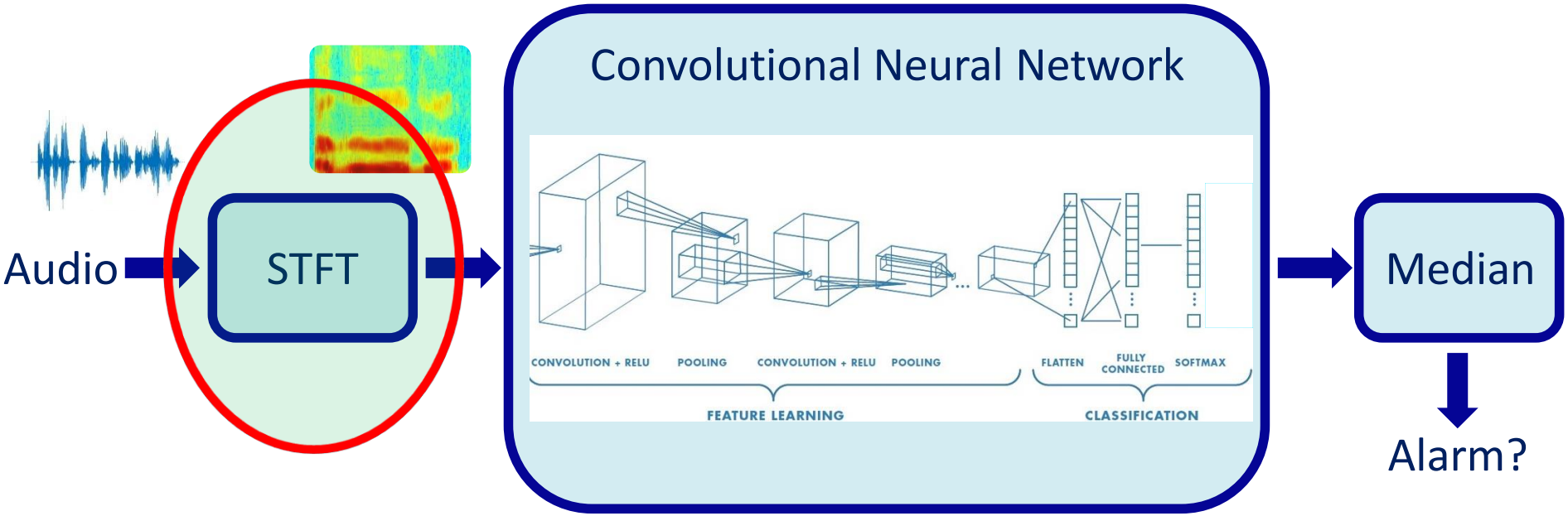
Detection of Alarm Sounds in Noisy Environments

Dean Carmel, Ariel Yeshurun, Yair Moshe
Signal and Image Processing Laboratory (SIPL)
Andrew and Erna Viterbi Faculty of Electrical Engineering, Technion – Israel Institute of Technology
Technion City, Haifa, Israel, http://sipl.technion.ac.il/

EUSIPCO 2017

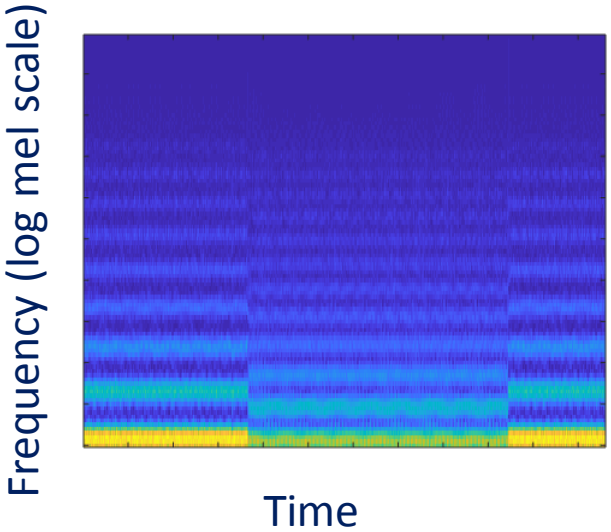
- Use classical techniques
  - A large set of acoustic features
    - Pitch, energy, MFCC, spectral features, wavelet coefficients, …
  - SVM classifier
- Results:
  - √ 98% accuracy with 200 msec delay
  - x Non-negligible false alarm rate
  - x Tested on a small dataset -> overfitting
    - 35 alarm sounds, 35 noises

Alarm Sounds Detection

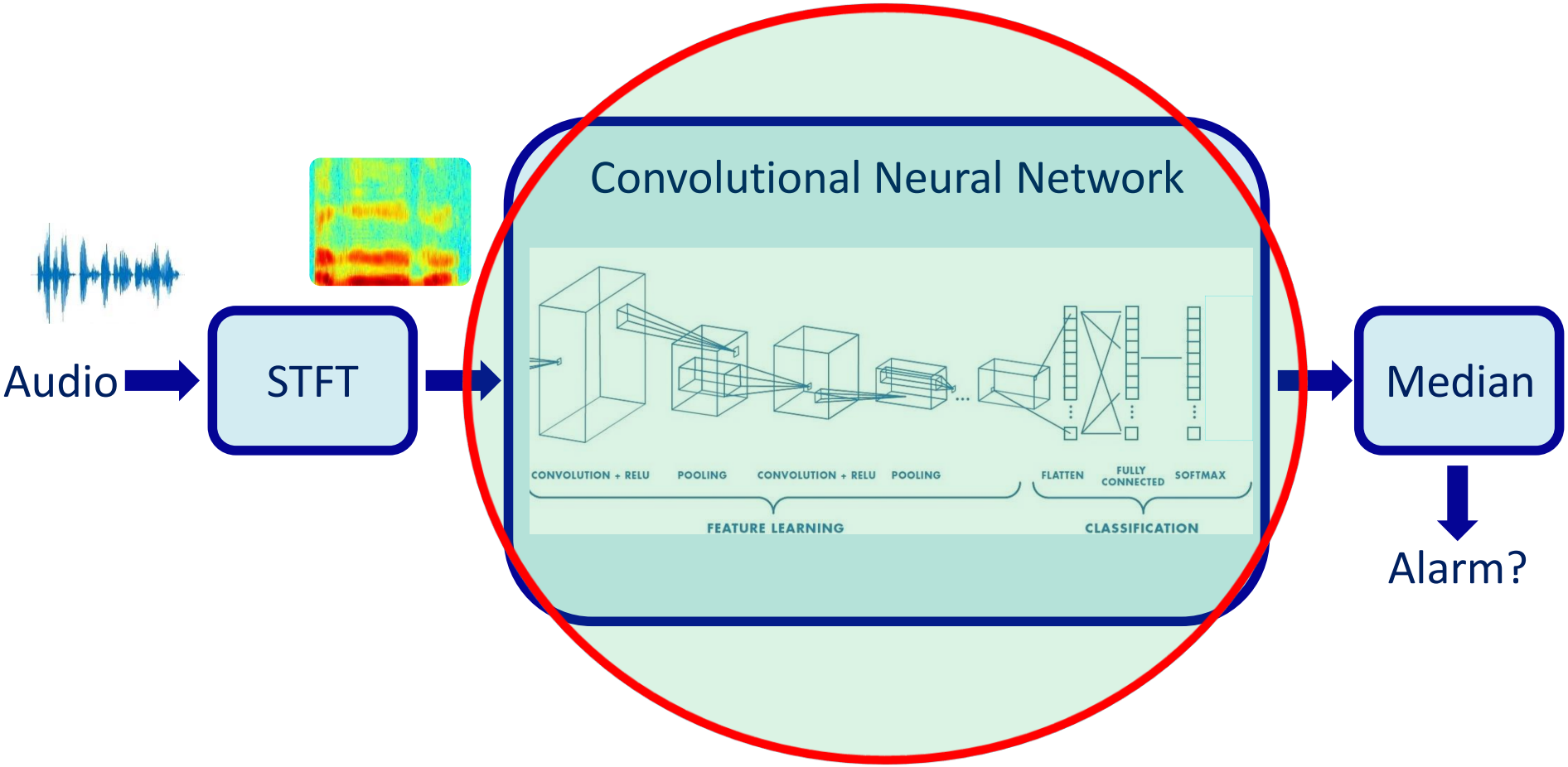


SFTF

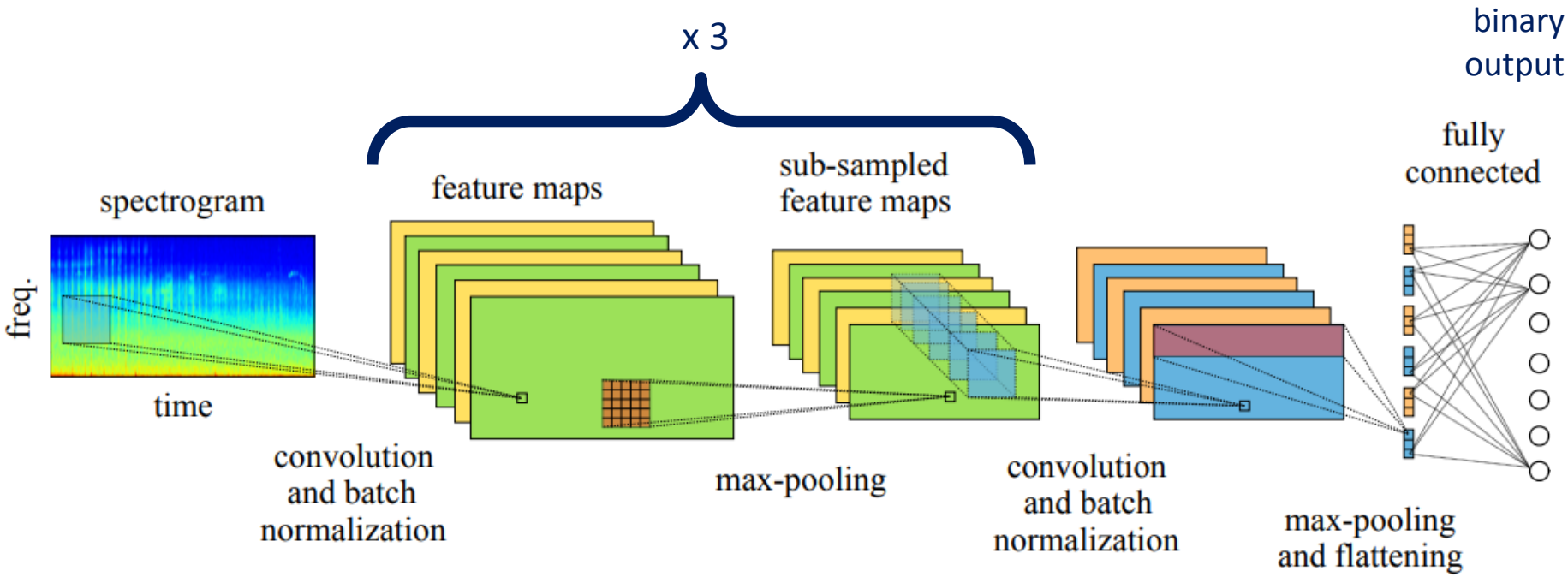
- Use log-scaled mel spectrograms
  - A perceptual frequency scale
  - Each ”pixel” in the spectrogram represents a frequency range
- Window size of 1 second
  - 50% overlap
  - 1 second → 580x100 “pixels”



Alarm Sounds Detection



CNN Architecture



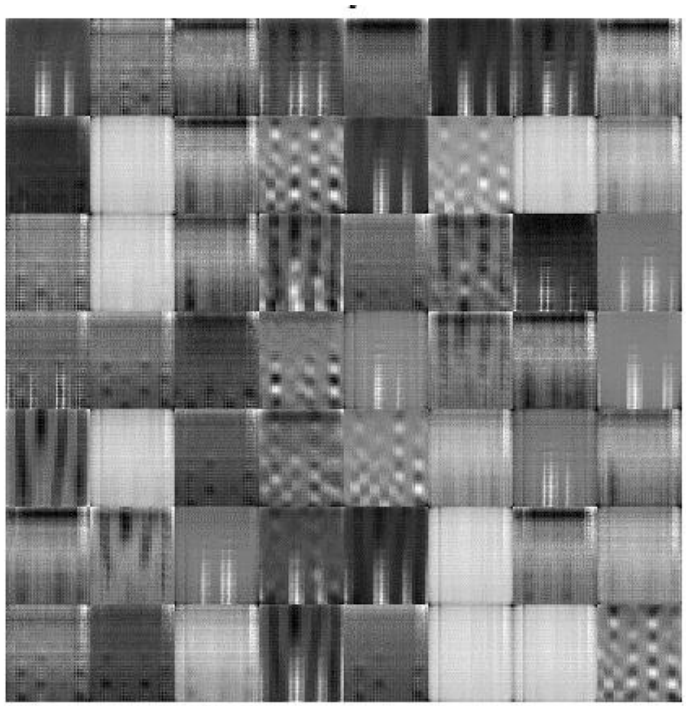
Initial Results

- On small dataset from our paper: Slightly better results than previous work
  - 35 alarm sounds, 35 noises

DeepDream

- Visualize image features learned by a network
  - Synthesizes images that strongly activate network layers
  - Available in MATLAB since R2017a

Last convolution layer



A Larger Dataset

- Urbansound8k dataset
  - 8k labeled urban sound excepts (≤4 sec) from 10 classes
  - 77 siren
- ESC-50 dataset
  - 2k labeled sound excepts (5 sec) from 50 classes
  - 39 sirens
- In Total:
  - 151 sirens
  - About 100,000 STFTs

- Intermediate results: 88% accuracy

Dataset Refinement

- Examples of problematic alarm sounds:

- Solution: manual labeling of each audio frame into one of several categories
  - Clean alarm
  - Noisy alarm
  - Distant alarm
  - No-alarm
  - Silence

Ongoing Work

- Parameter tuning with refined dataset
- More advanced architectures – ResNet, Inception, ..
- Unsupervised/semi-supervised learning

- Future: Implementing as an Android application

Lab Experiment in Our Faculty

- Lab experiments are performed by student pairs in their 3rd or 4th year of undergraduate studies
- Each student must select and carry out 7 out of 35 experiments
- Offered in different areas of electrical engineering
- Require only basic relevant theoretical knowledge
- Each experiment is divided into two sessions, 4-hours each

A New Lab Experiment

- ***SIPL*** is developing a new "deep learning" lab experiment
  - Covers basic machine/deep learning concepts
- Expected to be a very popular experiment

- Constraints:
  - No machine learning course as a prerequisite
  - Cannot assume knowledge of Python
  - Tight time constraints

- Solution: We are using MATLAB ☺

Just a Few Out of Many Activities

Thank You!

http://sipl.technion.ac.il/

https://yair.net.technion.ac.il/
yair@ee.technion.ac.il