INSTANT SEARCH - HSMW_TUC Team

TRECVID 2019

Tony Rolletschke - University of Applied Science Mittweida
Our INS approach with pretrained models and web based interactive evaluation by using a management system

Who is hiding behind ‘our’?

- Rico Thomanek
- Christian Roschke
- Benny Platte
- Tony Rolletschke
- Tobias Schlosser
- Manuel Heinzig
- Danny Kowerko
- Matthias Vodel
- Frank Zimmer
- Maximilian Eibl
- Marc Ritter
Our INS approach with pretrained models and web based interactive evaluation by using a management system

- We mainly wanted to improve and extend our system from last year
- While using the same hardware we could increase the system performance significantly
- That was the result of a new approach with distributed clusters administrated by a management system
- The focus was on high adaptability, which enables processing tasks in different application domains
Holistic server-client approach
Database
- Handles the persistent storage
- All data can be accessed by an API

Management
- Ad or remove processing nodes
- Tasks can allocated with different nodes
- Automatic task completion
- Error handling
System - Architecture

Distribution with docker

Person Recognition Unit
- Detectron
- YOLO9000
- Face Recognition

Activity Recognition Unit
- FaceNet
- OpenFace
- Face Recognition

Save size, position and class of object
Apply all result sets to any facial framework
Save recognized persons

Postprocessing / Scoring Unit

Management System

Keyframes

Synthetic Activity Data

Trainings Extension

Visualization Unit

Export Unit

Person References

Trainings Extension

Texting Phone
- Calling Phone
- Carrying
- Closing
- Entering Car
- Closing Trunk
- Opening
- Pick Up
- Pull

System - Architecture

Distribution with docker
System - Architecture

Distribution with docker

- Distribution without manual configuration
- Containers to be run multiple times on each node
- Each host can run on maximum load
State-of-the-art frameworks to identify people and objects:

- **System - Architecture**
- **Keyframes**
- **Synthetic Activity Data**
- **Trainings Extension**
- **Visualization Unit**
- **Export Unit**
- **Postprocessing / Scoring Unit**

**Person Recognition Unit**
- Detectron
- YOLO9000
- Face Recognition
- FaceNet
- OpenFace
- Face Recognition
- Save recognized persons

**Activity Recognition Unit**
- Apply all result sets to any facial framework
- Save size, position and class of object

**Management System**
- API

**References**
- Trainings
- Extension

**Texting Phone**
- Calling Phone
- Carrying
- Closing
- Entering Car
- Closing Trunk
- Opening
- Pick Up
- Pull
State-of-the-art frameworks to identify people and objects:

- Detectron
- Yolo9000
- FaceNet
- OpenFace
- FaceRecognition
- TuriCreate
Object detection with *Detectron*:

Real-time object detection with *Yolo9000*.

Face recognition with *OpenFace*:

Source: Amos, Brandon and Bartosz Ludwiczuk and Satyanarayanan, Mahadev, OpenFace: A general-purpose face recognition, 2016
Detect and recognize faces with *Face Recognition*:

Source: https://github.com/ageitgey/face_recognition
Simplify the development of custom machine learning models with Turi Create:

<table>
<thead>
<tr>
<th>ML Task</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recommender</td>
<td>Personalize choices for users</td>
</tr>
<tr>
<td>Image Classification</td>
<td>Label images</td>
</tr>
<tr>
<td>Drawing Classification</td>
<td>Recognize Pencil/Touch Drawings and Gestures</td>
</tr>
<tr>
<td>Sound Classification</td>
<td>Classify sounds</td>
</tr>
<tr>
<td><strong>Object Detection</strong></td>
<td>Recognize objects within images</td>
</tr>
<tr>
<td>One Shot Object Detection</td>
<td>Recognize 2D objects within images using a single example</td>
</tr>
<tr>
<td>Style Transfer</td>
<td>Stylize images</td>
</tr>
<tr>
<td><strong>Activity Classification</strong></td>
<td>Detect an activity using sensors</td>
</tr>
<tr>
<td>Image Similarity</td>
<td>Find similar images</td>
</tr>
<tr>
<td>Classifiers</td>
<td>Predict a label</td>
</tr>
<tr>
<td>Regression</td>
<td>Predict numeric values</td>
</tr>
<tr>
<td>Clustering</td>
<td>Group similar datapoints together</td>
</tr>
<tr>
<td>Text Classifier</td>
<td>Analyze sentiment of messages</td>
</tr>
</tbody>
</table>

Source: https://github.com/apple/turicreate
Workflow - Person Recognition Unit

Keyframes

Person Recognition Unit
- Detectron
- YOLO9000
- Face Recognition

Save size, position and class of object
- Apply all result sets to any facial framework
- FaceNet
- OpenFace
- Face Recognition

Save recognized persons

Activity Recognition Unit

Synthetic Activity Data

Trainings Extension

Export Unit

Visualization Unit
- XML
- JSON

Postprocessing / Scoring Unit

Management System

API

Person References

Trainings Extension
Person recognition in three steps

- Localization of the absolute position
  - Detectron 2.515.332 person objects
  - Face Recognition 1.384.747 faces
  - YOLO9000 1.013.007 person objects
- Each results stored separate in database table
Person recognition in three steps

- Localization of the absolute position
  - Detectron 2.515.332 person objects
  - Face Recognition 1.384.747 faces
  - YOLO9000 1.013.007 person objects
  - Each results stored separate in database table

- Person recognition
  - Classification of the persons
  - Usage of models trained on dataset collected with google-image-downloader
Person recognition in three steps

- Create ground truth data
  - Usage of google-image-downloader
  - Collection an average of 150 images/person
- Cut out faces with *Face Recognition*
- Train person models with three frameworks

```
$ googleimagesdownload -k "eastenders Mark Wingett" -s ">1024*768"
```

Google images download as a Docker container

Automatic removal of wrong person images

Grout Truth for a specific person
Person recognition in three steps

- Merging the results of all frameworks to a person score

\[
\text{predPerson} = \sum_{h=0}^{r} (x_h \in K),
\]

\[K = \{ x | 0 \leq x \leq 100 \},\]

\[\text{predPerson} \in L,\]

\[L = \{ y | 0 \leq y \leq 900 \}.
\]
Scoring Results

- We developed 2018 a web-service for interactive evaluation
- The findings from the visual processing not include in the automatic evaluation
- Number of fail detections decrease with increasing sore

<table>
<thead>
<tr>
<th>Score</th>
<th>Minty</th>
<th>Max</th>
<th>Jack</th>
<th>Heather</th>
<th>Zainab</th>
<th>Darrin</th>
<th>Mo</th>
<th>Garry</th>
<th>Jane</th>
<th>Chelsea</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>100</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>150</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>200</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>250</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>300</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>350</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>400</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>450</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>500</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>550</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>600</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>650</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>700</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>750</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>800</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Number of person recognition:

- Median: 105.040
- Maximum: 121.910
Workflow - Activity Recognition Unit

Different approaches for activity recognition

- Self-developed activity classifier
  - Create our own suitable synthetic ground truth dataset
  - Extracting body-key-points based on the results by OpenPose
  - Able to predict activities in realtime
Different approaches for activity recognition

- **Object Activity classifier (Detectron, YOLO9000, Custom-Model)**
  - The classifier was trained with 9504 images showing drinking, eating, holdingBaby, holdingGlass, holdingPhone, hugging, laughing and kissing
  - Images are downloaded with *GoogleDownloader* and manually labelled using *RectLabel* software
  - The object recognition classify and localize the activities

Source: Screenshot macOS app „RectLabel“
Fusion of the determined scoring values

- The fusion took place exclusively in the database
- Linked the various framework results with suitable SQL statements
- An overall score is be calculated with the person and activity recognition results
Our INS approach with pretrained models and web based interactive evaluation by using a management system

- The system compute twelve times more frames as last year in the same time
- So we significantly improved performance
- The using of the distributed approach lets us calculate parallel on serval machines
- A functional optimization of the newly developed parts follows next year
- So an increase in result quality after refinement is expected
- This was a practical evidence, that our optimized application is able to handle the provided data in an acceptable computational time
- With the focus on high adaptability the system could be extended every time by frameworks of other application domains