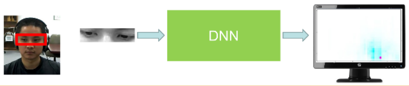
Estimation of Gaze Region using Two Dimensional Probabilistic Maps Constructed using Convolutional Neural Networks

Sumit Jha, Carlos Busso

Motivation

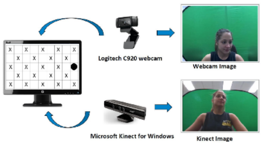
Background:

- Gaze tracking can be helpful in understanding user's engagement
  - Student's attention in remote learning
  - Distraction during driving
  - Interaction in human-robot and human-computer interfaces
- Target system: Calibration-free gaze estimation
- Approach
  - Predicting a probabilistic confidence region
  - Solving regression as a classification task
  - CNN: downsampling followed by upsampling



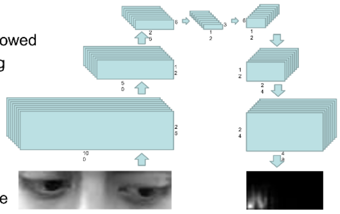
MSP-GAZE

- Gaze corpus collected at UT-Dallas [Li,2018]
- Target point projected on the highlighted portion of the monitor
- Data collected with 46 subjects
  - Gender balanced
  - Diverse ethnic group
  - Multiple sessions
- RGB data from the webcam is used
- Eye pair obtained using Viola-Jones algorithm
  - Grayscale
  - Resized to 100 x 25



Model

- Network purely based on convolutional layers
- Sequence of max-pooling followed by a sequence of up sampling
- Output is obtained as a label in the grid
- 16, 3x3 filters at each stage
- ReLU activation
- Input – 100x25 eye pair image
- Output – 48x24 grid
- Subject independent partition



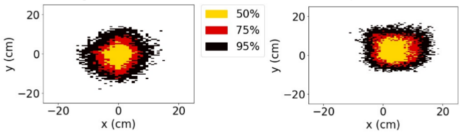
- Output resolution can be adjusted based on application by adding or removing layers
- Softmax activation at the last layer to output probability scores for each grid
- Cross entropy loss on weighted output to penalize larger error

Results

Comparison with Regression Model

- The predicted label is the output label with the highest value
- Baseline: regression model with similar architecture
  - 6 convolution layer followed by 2 fully connected layers
- More parameters in the regression model because of fully connected layers

Histogram of difference between true label and the predicted label



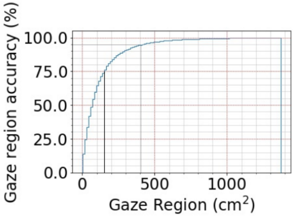
Proposed Approach

Regression model

Evaluation of Probabilistic Gaze Map

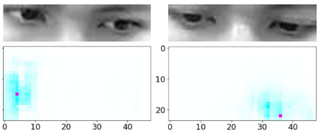
Accuracy versus resolution

- Confidence region with different resolution
- Larger areas – lower resolution, higher accuracy
- 75% accuracy at 13cm x 13cm



Probabilistic Map

- Distribution of gaze as softmax output
- More practical than deterministic output



Conclusions

- Probabilistic confidence region of gaze provides a practical method to estimate visual attention
- Easy integration with current models by replacing the fully connected layers with CNNs
- Less number of parameters and efficient implementation by multi-threading the code

Future Work

- Application in naturalistic driving condition
- More robust models
  - Lateral connections to maintain high spatial resolution
  - Ladder connections for semi-supervised learning

References:
Nanxiang Li and Carlos Busso, "Calibration free, user independent gaze estimation with tensor analysis," Image and Vision Computing, vol. 74, pp. 10-20, June 2018.

Semiconductor Research Corporation

This work was supported by Semiconductor Research Corporation (SRC) / Texas Analog Center of Excellence (TxACE), under task 2810.014

TxACE TEXAS ANALOG CENTER OF EXCELLENCE