**Objective**

Can saliency maps alone be used to classify task?

Task classification via eye movements (EMs) may be useful for:
- Hands-free computer operation
- Eye-movement gesture recognition
- Clinical disorder diagnosis (ADHD, Alzheimer’s)

Determination of task typically achieved using aggregate EM measures, scanpaths, cluster analysis. These require significant preprocessing of data.

Analysis via raw EM patterns (i.e. saliency maps) can benefit:
- Adaptive computing (web search, marketing)
- Onboard analysis in glasses or HUDs
- Biometrics for identification and classification

---

**Data Collection**

52 participants, within-subjects design
14 images per task, per subject (1456 total samples)

Search task = “Count the number of animals”
Judgement task = “How aesthetically pleasing is this?”

Monochrome saliency maps were generated for each sample from raw X,Y coordinates.

---

**Analysis and Results**

Capsule Networks analysis of SalMaps resulted in significantly improved classification accuracy, loss:

Traditional CNN (AlexNet):
- 4 Convolutional/Pooling Layers
- 3 Fully Connected Layers (with 20% dropout)
- 53.73% accuracy (70.11% loss)
- Performs at chance

Capsule CNN:
- 4 Convolutional layers
- 2 Capsule layers (1 Squash, 1 16D Routing)
- 3 Fully Connected Layers (with 20% dropout)
- 89.71% accuracy (14.70% loss)

Train/Valid/Test split at 60/20/20
Models ran 20 times with random splitting, accuracy averaged with p <.05

---

**Overview of Capsule Networks**

Traditional CNNs use pooling (segmentation) to reduce processing power. Each segment is processed individually.

Capsule Networks use capsules to group these segments via a progressively specialized hierarchy of convolutions.

CapsNets utilize routing to process image features (i.e. eyes) in separate capsules, followed by squashing to then process the whole image based on capsule output.

---

**Aggregate EM Metric Analysis**

Use of classification coefficient with aggregate EM metrics results in significant improvement of accuracy for task identification.

Aggregate EM metrics: Total Fixation Duration (TFD) Fixation Count (FC)

Logistic regression with:
- only TFD + FC = 74.48% acc.
- only ClassCoef = 91.71% acc.
- both TFD/FC and ClassCoef = 95.25% acc.

Multilayer Perceptron (4 layers) = 97.07% acc. (9.30% loss)