FEW-SHOT PERSONALIZED SALIENCY PREDICTION USING PERSON SIMILARITY BASED ON COLLABORATIVE MULTI-OUTPUT GAUSSIAN PROCESS REGRESSION

【PAPER ID: 3136】

Yuya Moroto, Keisuke Maeda, Takahiro Ogawa and Miki Haseyama *Hokkaido University, Japan*
E-mail: moroto@lmd.ist.hokudai.ac.jp

IEEE ICIP 2021

INTRODUCTION

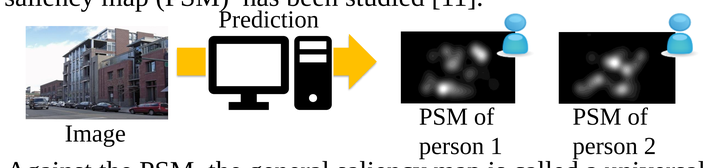
Saliency map

The saliency map represents an attractive image component in an image by imitating the instinctive human visual attention [3].

✘ However, the actual gazed area can differ from person to person.

Personalized saliency map

In order to predict the actual gazed area in the image, the personalized saliency map (PSM) has been studied [11].



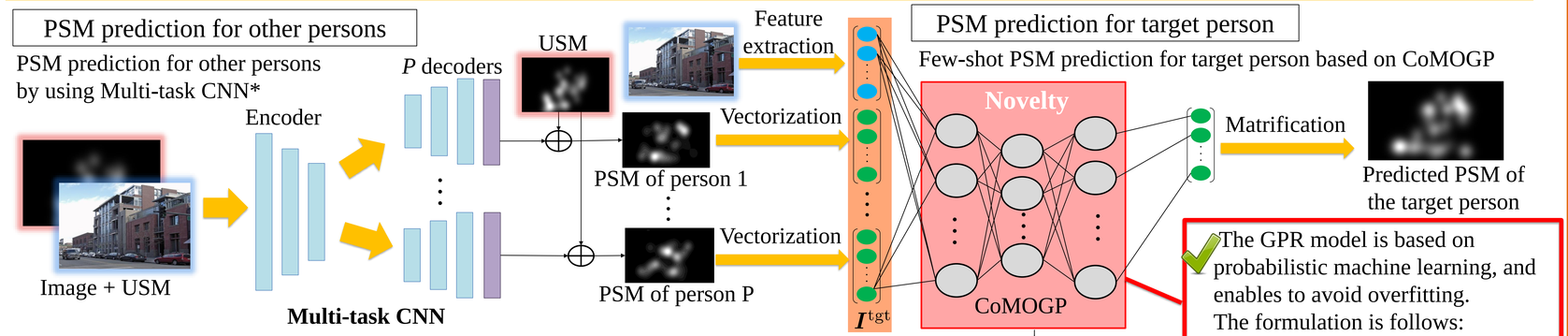
Against the PSM, the general saliency map is called a universal saliency map (USM).

✘ Although gaze data should be obtained from the person for the PSM prediction, the acquisition of those data is the burden on the person.
➤ The deterministic machine learning method can overfit the small amount of gaze data.

By using probabilistic machine learning and gaze data obtained from other persons, the few-shot PSM prediction is expected to realize.

PROPOSED METHOD

Few-shot PSM Prediction Based on Collaborative Multi-output Gaussian Process Regression (CoMOGP) [19]



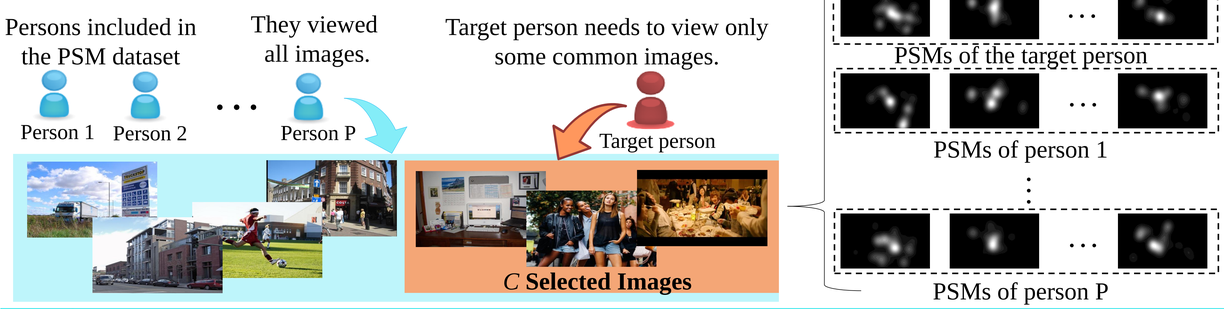
The GPR model is based on probabilistic machine learning, and enables to avoid overfitting.
The formulation is follows:

$$p(\boldsymbol{g}(\boldsymbol{I}^{\mathrm{tgt}}))|\boldsymbol{I},\boldsymbol{Y},\boldsymbol{I}^{\mathrm{tgt}}) \sim \mathcal{N}(\boldsymbol{\mu}^{\mathrm{tgt}},\boldsymbol{\sigma}^{\mathrm{tgt}})$$

$\boldsymbol{I}^{\mathrm{tgt}}$: Input of the target sample
$\boldsymbol{I}$: Inputs of training samples
$\boldsymbol{Y}$: PSMs of training images
$g(\cdot)$: Output from CoMOGP
$\boldsymbol{\mu}^{\mathrm{tgt}}$: Mean of trained distribution
$\boldsymbol{\sigma}^{\mathrm{tgt}}$: Variance of trained distribution

Adaptive image selection (AIS) [18] for training CoMOGP

Images that should be gazed at are selected based on the variance in PSMs of objects.



By using the gaze data obtained from other persons, our method complements the lack of data.

✓ The use of the probabilistic regression model is expected to reach the accurate few-shot PSM prediction without overfitting.

EXPERIMENTAL RESULTS

Dataset

PSM dataset [11]
- Num. images: 1600, Num. Subjects: 30
- Subjects gazed at each image for 3 seconds under the free-viewing condition.

Settings

Images were randomly separated as follows:
- Num. training: 1100, Num. test: 500

Subjects were randomly separated as follows:
- Num. target persons: 10

Num. selected images *C:* 100

Ground Truth

PSM calculated from gaze data by following [23]

Evaluation Index

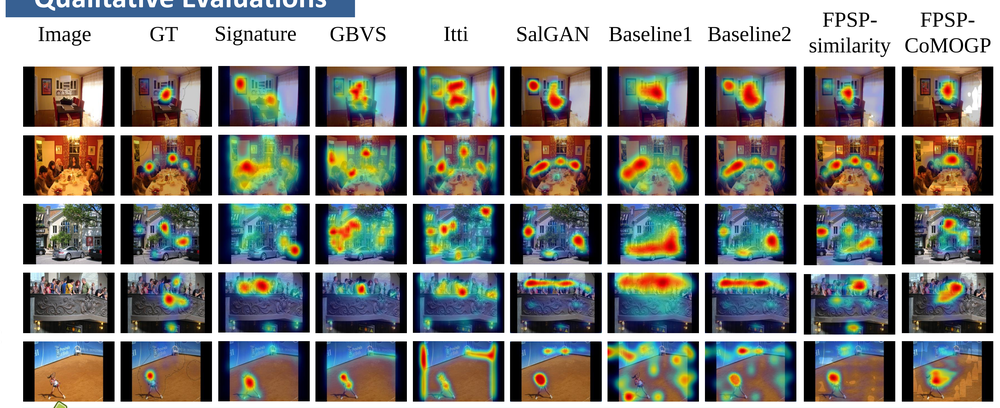
By referring to [26], we selected 3 metrics as follows:
- Pearson's correlation coefficient (CC) ↑
- histogram intersection (Sim) ↑
- Kullback-Leibler divergence (KLdiv) ↓

Comparison methods

Several USM prediction from MIT saliency benchmark [27]
- Signature [1], GBVS [2], Itti [3], SalGAN[4]

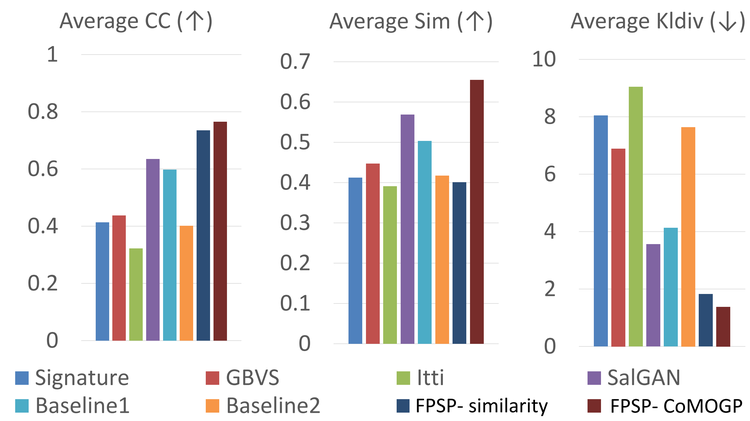
3 few-shot PSM prediction methods
- Baseline1 [29], Baseline2 [30], FPSP-similarity [18]

Qualitative Evaluations



✓ Visually confirmed that ours can predict PSM more accurately than others.

Quantitative Evaluations



FPSP-CoMOGP vs. Signature, GBVS, Itti, SalGAN
➤Confirmed the effectiveness of the person-specific prediction.

FPSP-CoMOGP vs. Baseline1, 2, FPSP-similarity
➤Confirmed the effectiveness of the GPR model.

✓ Verified that our method was effective for the few-shot PSM prediction.

Acknowledgement : This work was partly supported by the KAKENHI Grant Number JP17H01744.

*Multi-task Convolutional Neural Network