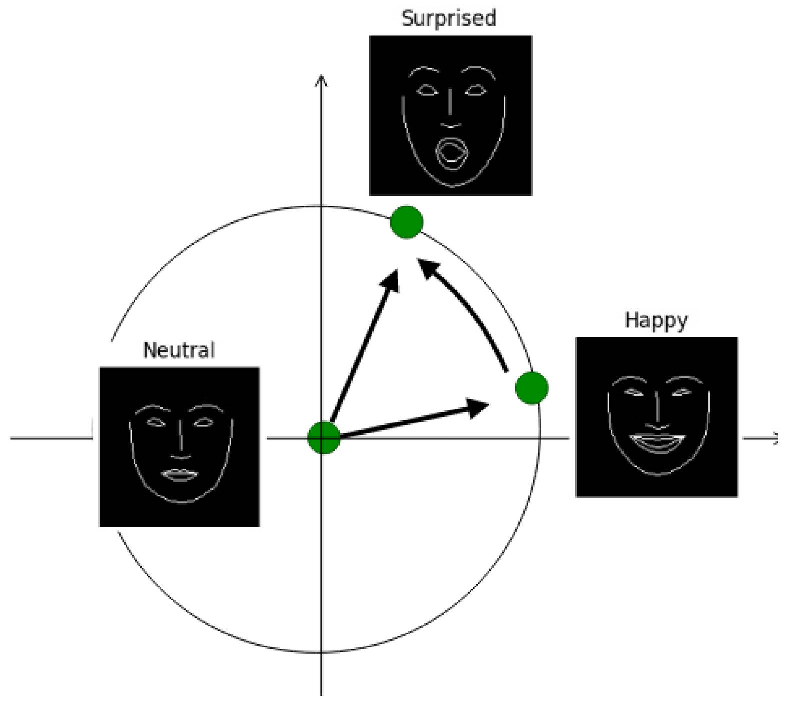
Continuous Emotion Transfer Using Kernels

A. Lambert†, S. Parekh†, Z. Szabó⋆, F. d'Alché-Buc†
†LTCI, Télécom Paris. ⋆Department of Statistics, London School of Economics.

Goal

Transfer an emotion to an input with continuous control



Input space $\mathcal{X} = \mathbb{R}^d$, emotion space $\Theta \subset \mathbb{R}^p$. Build

$$h : \mathcal{X} \times \Theta \mapsto \mathcal{X}, \text{ or equivalently } h : \mathcal{X} \mapsto (\Theta \mapsto \mathcal{X})$$

Empirical Risk $\mathcal{R}_\mathcal{S}(h)$

Training samples $(x_{ij}, \theta_{ij}^{\text{out}}, y_{ij})_{i\in[n], j\in S_i}$

- $x_{ij} \in \mathcal{X}$ corresponds to an object with input style $\theta_{ij}^{\text{in}} \in \Theta$
- $y_{ij} \in \mathcal{X}$ is the same object with output style $\theta_{ij}^{\text{out}} \in \Theta$
- For each object $i \in [n]$ we have access to $|S_i|$ style transition pairs $\left\{(\theta_{ij}^{\text{in}}, \theta_{ij}^{\text{out}})\right\}_{j\in S_i}$
- $\ell = \frac{1}{2}\|\cdot\|_\mathcal{X}^2$ is the square loss

$$\mathcal{R}_\mathcal{S}(h) := \frac{1}{n}\sum_{i\in[n]} \frac{1}{|S_i|}\sum_{j\in S_i} \ell\Big(\overbrace{h\big(\underbrace{x_{ij}}_{\text{input object}}\big)\big(\underbrace{\theta_{ij}^{\text{out}}}_{\text{output style}}\big)}^{\text{predicted output object}}, \underbrace{y_{ij}}_{\text{output object}}\Big)$$

Problem Formulation in vv-RKHSs

Extension of kernel methods to handle vector-valued outputs [1].

- $k_\mathcal{X} : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ and $k_\Theta : \Theta \times \Theta \to \mathbb{R}$ two scalar-valued kernels
- $\mathbf{A} \in \mathbb{R}^{d\times d}$ encoding similarities in the outputs
- $G = k_\Theta \mathbf{A}$, $K = k_\mathcal{X} \text{Id}_{\mathcal{H}_G}$ so that $h : \underbrace{\mathcal{X} \mapsto \overbrace{(\Theta \mapsto \mathcal{X})}^{\in \mathcal{H}_G}}_{\in \mathcal{H}_K}$

$$\min_{h\in\mathcal{H}_K} \mathcal{R}_\mathcal{S}(h) + \frac{\lambda}{2}\|h\|_{\mathcal{H}_K}^2, \quad \lambda > 0 \qquad (1)$$

Optimization

Lemma 1 (Representer) *Problem (1) has a unique solution $\hat{h}$ and it takes the form*

$$\hat{h}(x)(\theta) = \sum_{i=1}^{t}\sum_{j=1}^{m} k_\mathcal{X}(x, x_i) k_\Theta(\theta, \theta_{ij}) \mathbf{A} \hat{c}_{ij}, \quad \forall (x,\theta) \in \mathcal{X} \times \Theta$$

for some coefficients $\hat{c}_{ij} \in \mathbb{R}^d$ with $i \in [t]$ and $j \in [m]$.

• t depends on the number of style transition pairs $\left\{(\theta_{ij}^{\text{in}}, \theta_{ij}^{\text{out}})\right\}_{j\in S_i}$.

(Almost) Closed-form Solution: Reshaping the coefficients in a matrix $\hat{\mathbf{C}} \in \mathbb{R}^{tm\times d}$ yields the Sylvester equation

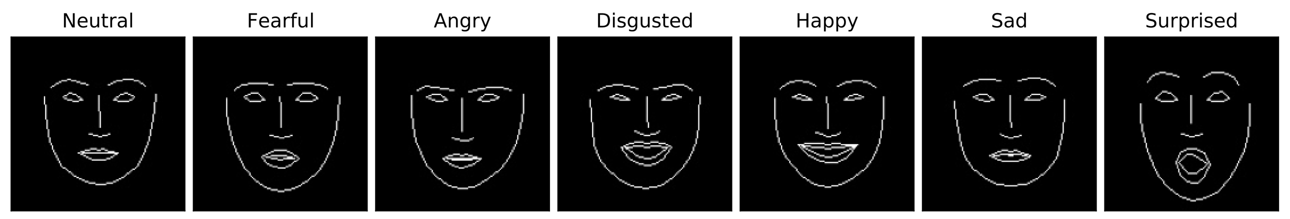
$$\mathbf{K}\hat{\mathbf{C}}\mathbf{A} + tm\lambda\hat{\mathbf{C}} = \mathbf{Y}$$

where $\mathbf{K} \in \mathbb{R}^{tm\times tm}$ is the gram matrix of the problem.
When $\mathbf{A} = \mathbf{I}_d$ (identity matrix of size $d \times d$), the solution is analytic:

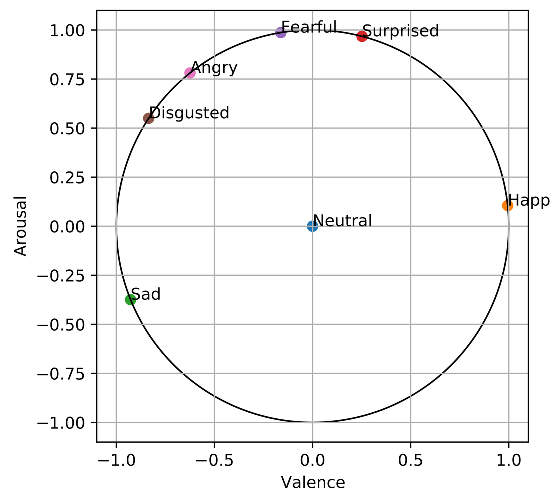
$$\hat{\mathbf{C}} = (\mathbf{K} + tm\lambda\mathbf{I}_{tm})^{-1}\mathbf{Y}.$$

Representation Choices

• Extract landmarks from face images (KDEF, RaFD datasets), $\mathcal{X} = \mathbb{R}^{136}$



• Represent emotions on a compact space $\Theta \subset \mathbb{R}^p$, e.g. *Valence-Arousal* representation when $p = 2$ [2].



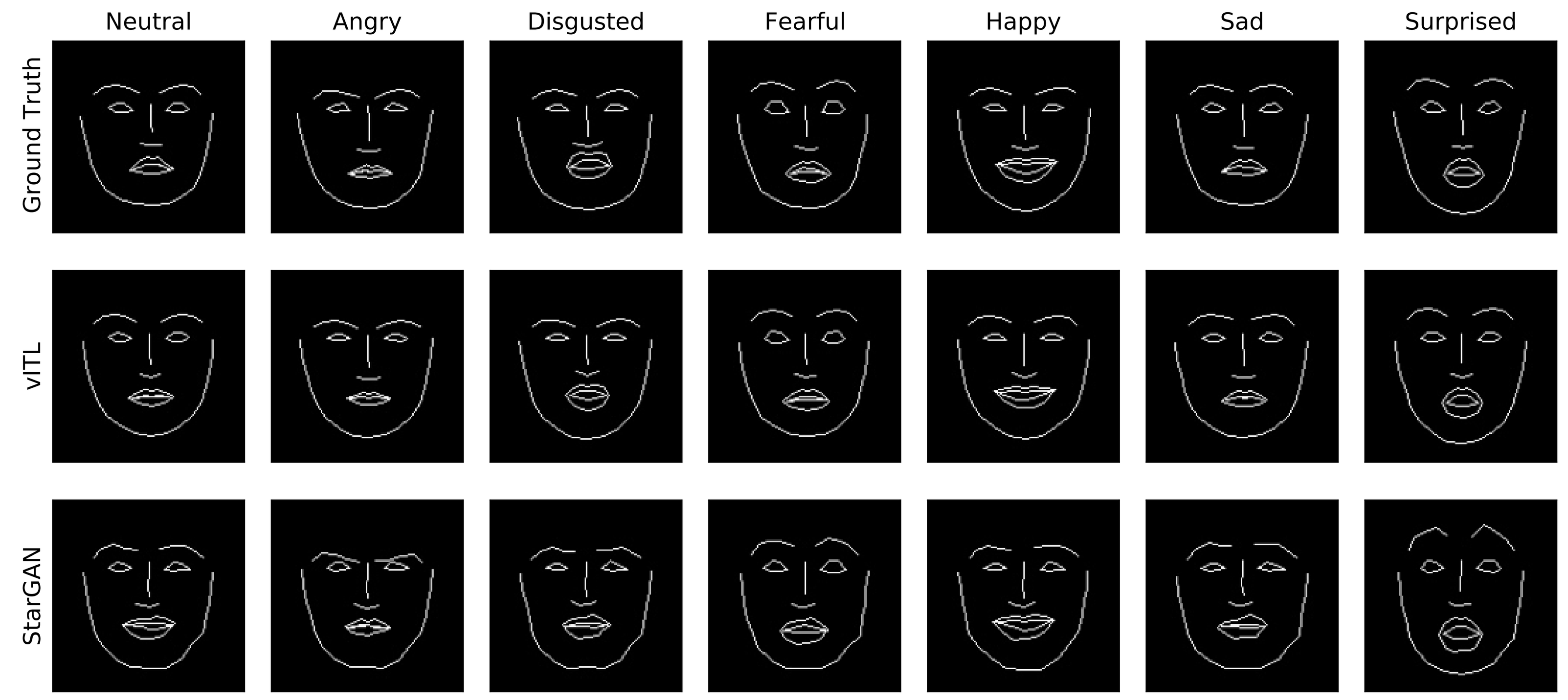
Quantitative Results

Comparison of vITL (ours) against StarGAN [3] in MSE on test set. Two scenarii depending on the observed θ^{in}: single emotional input θ_0 or joint model.

Experimental setup: $k_\mathcal{X}$, k_Θ are Gaussian kernels, $\mathbf{A} = \mathbf{I}_d$, kernel parameters and λ chosen through cross-validation.

Methods	KDEF frontal	RaFD frontal
vITL: θ_0 = neutral	0.010 ± 0.001	0.009 ± 0.004
vITL: θ_0 = fearful	0.010 ± 0.001	0.010 ± 0.005
vITL: θ_0 = angry	0.012 ± 0.002	0.010 ± 0.005
vITL: θ_0 = disgusted	0.012 ± 0.001	0.010 ± 0.004
vITL: θ_0 = happy	0.011 ± 0.001	0.010 ± 0.004
vITL: θ_0 = sad	0.011 ± 0.001	0.009 ± 0.004
vITL: θ_0 = surprised	0.010 ± 0.001	0.011 ± 0.006
vITL: Joint	**0.011** ± 0.001	**0.007** ± 0.001
Landmark-StarGAN	0.029 ± 0.003	0.024 ± 0.007

Qualitative Results



• Given a face x, able to generate the trajectory $\theta \to \hat{h}(x)(\theta)$.

Code Available

https://github.com/allambert/torch_itl

A.L. and S.P. were funded by the research chair Data Science & Artificial Intelligence for Digitalized Industry and Services at Télécom Paris.

Z.Sz. benefited from the support of the Europlace Institute of Finance and that of the Chair Stress Test, RISK Management and Financial Steering, led by the French École Polytechnique and its Foundation and sponsored by BNP Paribas.

[1] Carmeli, Claudio and De Vito, Ernesto and Toigo, Alessandro and Umanitá, Veronica Vector valued reproducing kernel Hilbert spaces and universality. In *Analysis and Applications*, vol 8 pp 19–61, 2010.
[2] Russell, James A A circumplex model of affect. In *Journal of Personality and Social Psychology*, pp. 1161–1178, 1980.
[3] Choi, Yunjey and Choi, Minje and Kim, Munyoung and Ha, Jung-Woo and Kim, Sunghun and Choo, Jaegul. StarGAN: Unified Generative Adversarial Networks for Multi-domain Image-to-Image Translation. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 8789–8797, 2018.