Continuous Emotion Transfer using RKHSs

Zoltán Szabó
Department of Statistics, LSE
(joint work with Alex Lambert, Sanjeel Parekh, Florence d'Alché-Buc)

LIKE
Jan. 12, 2022

Style transfer

- Goal: transfer an object according to a target style.

Numerous applications

- Computer vision [Ulyanov et al., 2016, Choi et al., 2018, Puy and Pérez, 2019, Yao et al., 2020], NLP [Fu et al., 2018], audio signal processing [Grinstein et al., 2018].
- Graphics: animating digital characters & avatars → body MOCAP [Aristidou et al., 2017, Aberman et al., 2020].

- Health & industry: digital twinning [Tao et al., 2019, Barricelli et al., 2019, Lim et al., 2020].

Style transfer

- Goal: transfer an object according to a target style.

Numerous applications

- Computer vision [Ulyanov et al., 2016, Choi et al., 2018, Puy and Pérez, 2019, Yao et al., 2020], NLP [Fu et al., 2018], audio signal processing [Grinstein et al., 2018].
- Graphics: animating digital characters & avatars → body MOCAP [Aristidou et al., 2017, Aberman et al., 2020].

- Health & industry: digital twinning [Tao et al., 2019, Barricelli et al., 2019, Lim et al., 2020].

Our aim: have a ' slider ' (in $\mathbb{R}^p$)

general principled style transfer: object → continuum of styles

One-page summary

Focus: novel task

continuous style transfer $\xleftarrow{\text{spec.}}$ functional output regression.

- Framework: vv-RKHS.
- Ingredients: similarity on
  - objects: $k_{\mathcal{X}}$,
  - style: k_{Θ},
  - continuous style space: $\Theta \subset \mathbb{R}^p$.
- Also handles: limited observation.
- Running example: emotion transfer.

Emotion transfer

- Given: set of emotions.
- Goal: transform object representations of
  - faces [Choi et al., 2018], hands [Irimia et al., 2019], body movement [Aristidou et al., 2017], ...
  - repr: 2D images, 3D meshes, body skeletons, MOCAP sequences.

Emotion transfer

- Given: set of emotions.
- Goal: transform object representations of
  - faces [Choi et al., 2018], hands [Irimia et al., 2019], body movement [Aristidou et al., 2017], . . .
  - repr: 2D images, 3D meshes, body skeletons, MOCAP sequences.

Our example

- style := emotion,
- object representation := facial landmark locations.

Emotion transfer

- Given: set of emotions.
- Goal: transform object representations of
  - faces [Choi et al., 2018], hands [Irimia et al., 2019], body movement [Aristidou et al., 2017], . . .
  - repr: 2D images, 3D meshes, body skeletons, MOCAP sequences.

Our example

- style := emotion,
- object representation := facial landmark locations.

Why facial landmarks?

- Useful for facial identification [Mane and Shah, 2019], expression analysis [Devries et al., 2014], medical diagnosis [Balaei et al., 2017].

Problem formulation

- Object space: $\mathcal{X}$. Style space: Θ.
- Goal: $\boxed{\text{(object, style)} \mapsto \text{object}}$, i.e. an

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

Problem formulation

- Object space: $\mathcal{X}$. Style space: Θ.
- Goal: (object, style) $\mapsto$ object, i.e. an

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

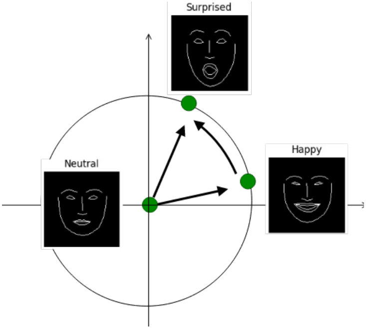
- In our case: landmarks $\mapsto$ $\underbrace{(\text{emotion} \mapsto \text{landmarks})}_{\text{function-valued regression}}$.

Problem formulation

- Object space: $\mathcal{X}$. Style space: Θ.
- Goal: $\boxed{\text{(object, style)} \mapsto \text{object}}$, i.e. an

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

- In our case: landmarks $\mapsto$ $\underbrace{(\text{emotion} \mapsto \text{landmarks})}_{\text{function-valued regression}}$.

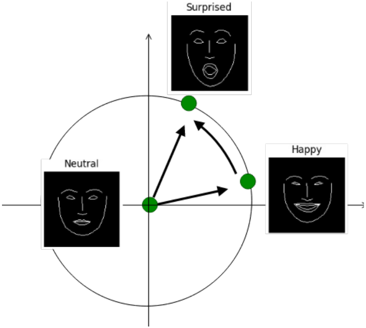


Problem formulation

- Object space: $\mathcal{X}$. Style space: Θ.
- Goal: $\boxed{\text{(object, style)} \mapsto \text{object}}$, i.e. an

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

- In our case: landmarks $\mapsto \underbrace{(\text{emotion} \mapsto \text{landmarks})}_{\text{function-valued regression}}$.



- $h : \mathcal{X} \mapsto (\Theta \mapsto \mathcal{Y})$ would work similarly [$\mathcal{Y}$ = avatars].

Cost function

- Training samples:
  - For each object $i \in [n]$: $|S_i|$ style transition pairs $\left\{\left(\theta^{\text{in}}_{i,j}, \theta^{\text{out}}_{i,j}\right)\right\}_{j \in S_i}$.
  - $x_{i,j} \in \mathcal{X}$: object with input style $\theta^{\text{in}}_{i,j}$,
  - $y_{i,j} \in \mathcal{X}$: object with output style $\theta^{\text{out}}_{i,j}$.

Cost function

- Training samples:
  - For each object $i \in [n]$: $|S_i|$ style transition pairs $\left\{\left(\theta_{i,j}^{\text{in}}, \theta_{i,j}^{\text{out}}\right)\right\}_{j \in S_i}$.
  - $x_{i,j} \in \mathcal{X}$: object with input style $\theta_{i,j}^{\text{in}}$,
  - $y_{i,j} \in \mathcal{X}$: object with output style $\theta_{i,j}^{\text{out}}$.
- Cost (quality of the reconstruction) of h:

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

Cost function

- Training samples:
  - For each object $i \in [n]$: $|S_i|$ style transition pairs $\left\{\left(\theta_{i,j}^{\text{in}}, \theta_{i,j}^{\text{out}}\right)\right\}_{j \in S_i}$.
  - $x_{i,j} \in \mathcal{X}$: object with input style $\theta_{i,j}^{\text{in}}$,
  - $y_{i,j} \in \mathcal{X}$: object with output style $\theta_{i,j}^{\text{out}}$.
- Cost (quality of the reconstruction) of h:

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

- Quadratic loss: $\ell = \frac{1}{2} \|\cdot\|_{\mathcal{X}}^2$.

Cost function

- Training samples:
  - For each object $i \in [n]$: $|S_i|$ style transition pairs $\left\{\left(\theta_{i,j}^{\text{in}}, \theta_{i,j}^{\text{out}}\right)\right\}_{j\in S_i}$.
  - $x_{i,j} \in \mathcal{X}$: object with input style $\theta_{i,j}^{\text{in}}$,
  - $y_{i,j} \in \mathcal{X}$: object with output style $\theta_{i,j}^{\text{out}}$.
- Cost (quality of the reconstruction) of h:

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

- Quadratic loss: $\ell = \frac{1}{2}\left\|\cdot\right\|_{\mathcal{X}}^2$.
- Hypothesis class (vv-RKHS): $\underbrace{h : \mathcal{X} \mapsto \underbrace{(\Theta \mapsto \mathcal{X})}_{\in\mathcal{F}:=\mathcal{H}_G}}_{\in\mathcal{H}:=\mathcal{H}_K}$.

Trajectory interpretation: elements of $\mathcal{F}$

- Person $i \in [n]$: captured by a trajectory $z_i \in \mathcal{F}$.

$$z_i: \text{ emotion } \theta \in \Theta \to \text{ landmarks } z_i(\theta) \in \mathcal{X}.$$

- z_i: observed at the style transition pairs $\left\{\left(\theta_{i,j}^{\text{in}}, \theta_{i,j}^{\text{out}}\right)\right\}_{j \in S_i}$.
- In other words,

$$x_{i,j} := z_i(\theta_{i,j}^{\text{in}}), \qquad y_{i,j} := z_i(\theta_{i,j}^{\text{out}}), \; i \in [n], j \in S_i.$$

Trajectory interpretation: elements of $\mathcal{F}$

- Person $i \in [n]$: captured by a trajectory $z_i \in \mathcal{F}$.

  $$z_i: \text{ emotion } \theta \in \Theta \to \text{ landmarks } z_i(\theta) \in \mathcal{X}.$$

- z_i: observed at the style transition pairs $\left\{\left(\theta_{i,j}^{\text{in}}, \theta_{i,j}^{\text{out}}\right)\right\}_{j \in S_i}$.
- In other words,

  $$x_{i,j} := z_i(\theta_{i,j}^{\text{in}}), \qquad y_{i,j} := z_i(\theta_{i,j}^{\text{out}}), \; i \in [n], j \in S_i.$$

- $h(x)(\theta)$: prediction from landmarks x and target emotion θ.

Emotion representation: $\Theta \subset \mathbb{R}^p$

- Classical categorical description:

  'happy', 'sad', 'angry', 'surprised', 'disgusted', 'fearful'.

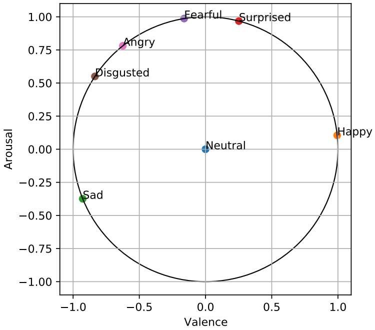
Emotion representation: $\Theta \subset \mathbb{R}^p$

- Classical categorical description:

  'happy', 'sad', 'angry', 'surprised', 'disgusted', 'fearful'.

- Valence-arousal model [Russell, 1980]: $\Theta \subset \mathbb{R}^2$,
  - valence: pleasure to displeasure,
  - arousal: high to low.

  Normalized demo:

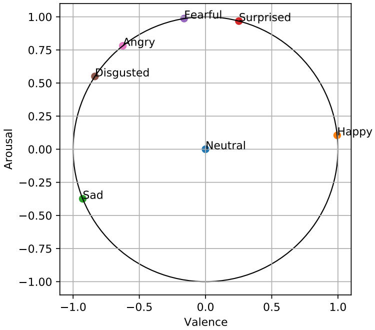


Emotion representation: $\Theta \subset \mathbb{R}^p$

- Classical categorical description:

  'happy', 'sad', 'angry', 'surprised', 'disgusted', 'fearful'.
- Valence-arousal model [Russell, 1980]: $\Theta \subset \mathbb{R}^2$,
  - valence: pleasure to displeasure,
  - arousal: high to low.

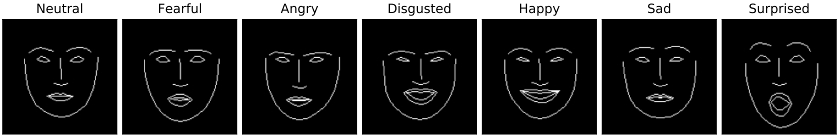
  Normalized demo:



- HigherD ($\Theta \subset \mathbb{R}^p$, $p \geq 2$) [Vemulapalli and Agarwala, 2019].

Object representation: $\mathcal{X} \subset \mathbb{R}^d$

- Face: landmarks points.
- Example: corners of the eyes, that of the mouth, ...



- Typically $M \approx 50 - 100 \Rightarrow$
  - $\mathcal{X} \subset \mathbb{R}^{d:=2M}$,
  - compact description – some trees: saved.

Two problem families

- Single emotional input:
  - input emotion: identical & fixed for everyone (θ_0).
  - output emotion: same m number.

$\Rightarrow$ I-O emotion pairs: $\{(\theta_0, \theta_{i,j})\}_{j \in [m]}$, $|S_i| = m \ \forall i$.

Two problem families

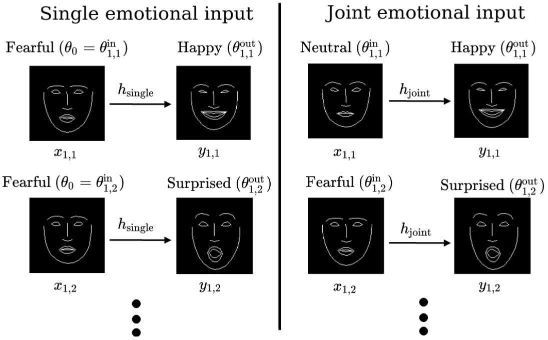
- Single emotional input:
  - input emotion: identical & fixed for everyone (θ_0).
  - output emotion: same m number.

  $\Rightarrow$ I-O emotion pairs: $\{(\theta_0, \theta_{i,j})\}_{j\in[m]}$, $|S_i| = m \; \forall i$.
- Joint emotional input:
  - m emotions for each person: $\{\theta_{i,a}\}_{a\in[m]}$, with all combinations,

  $\Rightarrow$ I-O emotion pairs: $\{(\theta_{i,a}, \theta_{i,b})\}_{a,b\in[m]}$, $|S_i| = m^2 \; \forall i$.

Two problem families

- Single emotional input:
  $\Rightarrow$ I-O emotion pairs: $\{(\theta_0, \theta_{i,j})\}_{j\in[m]}$, $|S_i| = m \ \forall i$.
- Joint emotional input:
  $\Rightarrow$ I-O emotion pairs: $\{(\theta_{i,a}, \theta_{i,b})\}_{a,b\in[m]}$, $|S_i| = m^2 \ \forall i$.



Towards the hypothesis class: $\mathcal{F}$

- Recall: $h : \mathcal{X} \mapsto \underbrace{(\underbrace{\boxed{\Theta \mapsto \mathcal{X}}}_{\in \mathcal{F} := \mathcal{H}_G})}_{\in \mathcal{H} := \mathcal{H}_K}$.

Towards the hypothesis class: $\mathcal{F}$

- Recall: $h : \mathcal{X} \mapsto \underbrace{(\underbrace{\boxed{\Theta \mapsto \mathcal{X}}}_{\in \mathcal{F} := \mathcal{H}_G})}_{\in \mathcal{H} := \mathcal{H}_K}$.
- $G : \Theta \times \Theta \to \underbrace{\mathcal{L}(\mathbb{R}^d)}_{\mathbb{R}^{d \times d}}$: matrix-valued kernel on Θ:
  1. $G(\theta, \theta') = G(\theta', \theta)^\top$ – symmetric,
  2. $\sum_{i,j \in [n]} v_i^\top G(\theta_i, \theta_j) v_j \geq 0$ – positive definite.

Towards the hypothesis class: $\mathcal{F}$

- Recall: $h : \underbrace{\mathcal{X} \mapsto \underbrace{(\boxed{\Theta \mapsto \mathcal{X}})}_{\in \mathcal{F} := \mathcal{H}_G}}_{\in \mathcal{H} := \mathcal{H}_K}$.
- $G : \Theta \times \Theta \to \underbrace{\mathcal{L}(\mathbb{R}^d)}_{\mathbb{R}^{d \times d}}$: matrix-valued kernel on Θ:
  1. $G(\theta, \theta') = G(\theta', \theta)^\top$ – symmetric,
  2. $\sum_{i,j \in [n]} v_i^\top G(\theta_i, \theta_j) v_j \geq 0$ – positive definite.
- Associated vv-RKHS: $\mathcal{H}_G = \overline{\mathrm{Span}} \left\{ G(\cdot, \theta)x : (\theta, x) \in \Theta \times \mathbb{R}^d \right\}$.

Towards the hypothesis class: $\mathcal{F}$ – continued

- Simple & popular: separable kernel

$$G(\theta,\theta') = k_{\Theta}(\theta,\theta')\mathbf{A},$$

kernel $k_{\Theta} : \Theta \times \Theta \to \mathbb{R}$, $\mathbf{0} \preccurlyeq \mathbf{A} \in \mathbb{R}^{d\times d}$.

Towards the hypothesis class: $\mathcal{F}$ – continued

- Simple & popular: separable kernel

$$G(\theta, \theta') = k_\Theta(\theta, \theta')\mathbf{A},$$

kernel $k_\Theta : \Theta \times \Theta \to \mathbb{R}$, $\mathbf{0} \preccurlyeq \mathbf{A} \in \mathbb{R}^{d \times d}$.

- Intuition:
  - smoothness: Gaussian kernel $k_\Theta(\theta, \theta') = \exp\left(-\gamma \|\theta - \theta'\|_2^2\right)$, $\gamma > 0$,
  - dependency among output coordinates: $\mathbf{A}$.

Hypothesis class: $\mathcal{H}$

- Recall: $h : \mathcal{X} \mapsto \underbrace{\underbrace{(\Theta \mapsto \mathcal{X})}_{\in \mathcal{F} := \mathcal{H}_G}}_{\in \boxed{\mathcal{H} := \mathcal{H}_K}}$.
- $K : \mathcal{X} \times \mathcal{X} \to \mathcal{L}(\mathcal{H}_G)$: operator-valued kernel:
  1. $K(x, x') = K(x', x)^*$ for all $x, x' \in \mathcal{X}$ – symmetric,
  2. $\sum_{i,j \in [n]} \langle f_i, K(x_i, x_j) f_j \rangle_{\mathcal{H}_G} \geq 0$ – positive definite.

Hypothesis class: $\mathcal{H}$ – continued

- Associated vv-RKHS:

$$\mathcal{H}_K = \overline{\mathrm{Span}} \left\{ K(\cdot, x) f : (x, f) \in \mathcal{X} \times \mathcal{H}_G \right\}.$$

Hypothesis class: $\mathcal{H}$ – continued

- Associated vv-RKHS:

$$\mathcal{H}_K = \overline{\mathrm{Span}}\left\{K(\cdot, x)f : (x, f) \in \mathcal{X} \times \mathcal{H}_G\right\}.$$

- Separable kernel:

$$K(x, x') = k_{\mathcal{X}}(x, x')\mathrm{Id}_{\mathcal{H}_G},$$

with kernel $k_{\mathcal{X}} : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$.

Hypothesis class: $\mathcal{H}$ – continued

- Associated vv-RKHS:

$$\mathcal{H}_K = \overline{\mathrm{Span}}\left\{K(\cdot, x)f : (x, f) \in \mathcal{X} \times \mathcal{H}_G\right\}.$$

- Separable kernel:

$$K(x, x') = k_{\mathcal{X}}(x, x')\mathrm{Id}_{\mathcal{H}_G},$$

with kernel $k_{\mathcal{X}} : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$.

- Intuition: smoothness $\Leftarrow k_{\mathcal{X}}$: Gaussian on $\mathcal{X}$.

Until now

- Wanted: $h : \mathcal{X} \mapsto \underbrace{(\underbrace{\boxed{\Theta \mapsto \mathcal{X}}}_{\in \mathcal{F} := \mathcal{H}_G})}_{\in \mathcal{H} := \mathcal{H}_K}$.
- Model: vv-RKHSs,

$$G(\theta, \theta') = k_\Theta(\theta, \theta')\mathbf{A}, \qquad K(x, x') = k_\mathcal{X}(x, x')\mathrm{Id}_{\mathcal{H}_G},$$

Until now

- Wanted: $h : \mathcal{X} \mapsto \underbrace{\underbrace{(\boxed{\Theta \mapsto \mathcal{X}})}_{\in \mathcal{F} := \mathcal{H}_G}}_{\in \mathcal{H} := \mathcal{H}_K}$.
- Model: vv-RKHSs,

$$G(\theta, \theta') = k_\Theta(\theta, \theta')\mathbf{A}, \qquad K(x, x') = k_\mathcal{X}(x, x')\mathrm{Id}_{\mathcal{H}_G},$$

Assumptions [$h : \mathcal{X} \mapsto (\boxed{\Theta \mapsto \mathcal{Y}})$]

- input object space ($\mathcal{X}$), style space (Θ): kernel-enriched.
- output object space ($\mathcal{Y}$): Hilbert.

Until now

- Wanted: $h : \mathcal{X} \mapsto \underbrace{(\underbrace{\Theta \mapsto \mathcal{X}}_{\in \mathcal{F} := \mathcal{H}_G})}_{\in \mathcal{H} := \mathcal{H}_K}$.
- Model: vv-RKHSs,

$$G(\theta, \theta') = k_\Theta(\theta, \theta')\mathbf{A}, \qquad K(x, x') = k_\mathcal{X}(x, x')\mathrm{Id}_{\mathcal{H}_G},$$

Assumptions [$h : \mathcal{X} \mapsto (\Theta \mapsto \mathcal{Y})$]

- input object space ($\mathcal{X}$), style space (Θ): kernel-enriched.
- output object space ($\mathcal{Y}$): Hilbert.

Specific case: ITL ($\mathcal{Y} = \mathbb{R}$, [Brault et al., 2019])

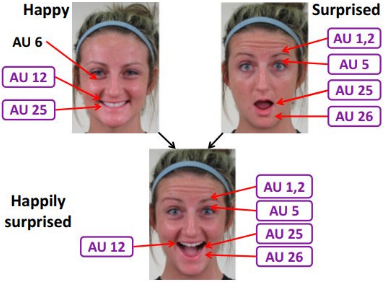
joint quantile regression, cost-sensitive classification, density level set estimation.

A few kernel domains ($\mathcal{X}$, Θ)

- Strings [Watkins, 1999, Lodhi et al., 2002, Leslie et al., 2002, Kuang et al., 2004, Leslie and Kuang, 2004, Saigo et al., 2004, Cuturi and Vert, 2005],
- time series [Rüping, 2001, Cuturi et al., 2007, Cuturi, 2011, Király and Oberhauser, 2019],
- trees [Collins and Duffy, 2001, Kashima and Koyanagi, 2002],
- groups and specifically rankings [Cuturi et al., 2005, Jiao and Vert, 2016],
- sets [Haussler, 1999, Gärtner et al., 2002], probability distributions [Berlinet and Thomas-Agnan, 2004, Hein and Bousquet, 2005, Smola et al., 2007, Sriperumbudur et al., 2010],
- various generative models [Jaakkola and Haussler, 1999, Tsuda et al., 2002, Seeger, 2002, Jebara et al., 2004],
- fuzzy domains [Guevara et al., 2017], or
- graphs [Kondor and Lafferty, 2002, Gärtner et al., 2003, Kashima et al., 2003, Borgwardt and Kriegel, 2005, Shervashidze et al., 2009, Vishwanathan et al., 2010, Kondor and Pan, 2016, Bai et al., 2020, Borgwardt et al., 2020].

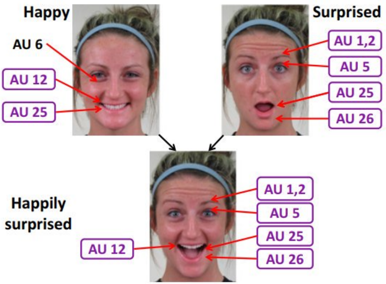
FACS and time series: emotion recognition

FACS (facial action coding system):



FACS and time series: emotion recognition

FACS (facial action coding system):

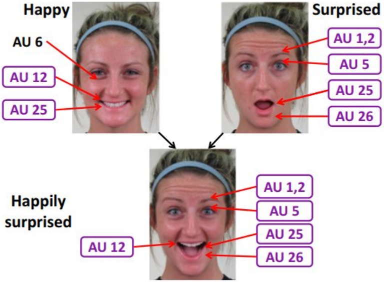


Idea:

- muscle activities $\mapsto$ emotion,
- landmarks $\approx$ simplified FACS,

FACS and time series: emotion recognition

FACS (facial action coding system):

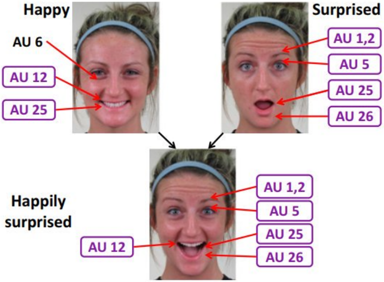


Idea:

- muscle activities $\mapsto$ emotion,
- landmarks $\approx$ simplified FACS,
- time series: can further increase accuracy,

FACS and time series: emotion recognition

FACS (facial action coding system):



Idea:

- muscle activities $\mapsto$ emotion,
- landmarks $\approx$ simplified FACS,
- time series: can further increase accuracy,
- Similarly: action recognition – writing, sports, games (Wii), ...

Optimization

- vv-RKHS $\Rightarrow$ rich still tractable, natural regularization.
- Task (vITL):

$$\min_{h \in \mathcal{H}_K} \mathcal{R}_\lambda(h) := \underbrace{\mathcal{R}_{\mathcal{S}}(h)}_{\text{data fitting}} + \frac{\lambda}{2} \underbrace{\|h\|^2_{\mathcal{H}_K}}_{\text{smoothness}}, \quad \lambda > 0.$$

Optimization

- vv-RKHS $\Rightarrow$ rich still tractable, natural regularization.
- Task (vITL):

$$\min_{h\in\mathcal{H}_K} \mathcal{R}_\lambda(h) := \underbrace{\mathcal{R}_{\mathcal{S}}(h)}_{\text{data fitting}} + \frac{\lambda}{2} \underbrace{\|h\|^2_{\mathcal{H}_K}}_{\text{smoothness}}, \quad \lambda > 0.$$

- Representer lemma:

$$\hat{h}(x)(\theta) = \sum_{i=1}^{t}\sum_{j=1}^{m} k_{\mathcal{X}}(x, x_i) k_{\Theta}(\theta, \theta_{i,j}) \mathbf{A}\hat{\mathbf{c}}_{i,j}, \ \{\hat{\mathbf{c}}_{i,j}\}_{i\in[t], j\in[m]} \subset \mathbb{R}^d.$$

Optimization

- vv-RKHS $\Rightarrow$ rich still tractable, natural regularization.
- Task (vITL):

$$\min_{h \in \mathcal{H}_K} \mathcal{R}_\lambda(h) := \underbrace{\mathcal{R}_{\mathcal{S}}(h)}_{\text{data fitting}} + \frac{\lambda}{2} \underbrace{\|h\|^2_{\mathcal{H}_K}}_{\text{smoothness}}, \quad \lambda > 0.$$

- Representer lemma:

$$\hat{h}(x)(\theta) = \sum_{i=1}^{t} \sum_{j=1}^{m} k_{\mathcal{X}}(x, x_i) k_{\Theta}(\theta, \theta_{i,j}) \mathbf{A} \hat{\mathbf{c}}_{i,j}, \quad \{\hat{\mathbf{c}}_{i,j}\}_{i \in [t], j \in [m]} \subset \mathbb{R}^d.$$

- With quadratic loss (ℓ): linear equation to $\hat{\mathbf{c}}_{i,j}$-s.

Quantitative illustration: setting

- 2 popular facial benchmarks: KDEF, RaFD.
- 68 2D landmarks: $M = 68, \mathbf{x} \in \mathbb{R}^{136=2\times 68}$.
- emotion representation: 2-dimensional valence-arousal ($\boldsymbol{\theta} \in \mathbb{R}^2$).
- $k_{\mathcal{X}}$, k_{Θ}: Gaussian kernels.

Quantitative illustration: setting

- 2 popular facial benchmarks: KDEF, RaFD.
- 68 2D landmarks: $M = 68, \mathbf{x} \in \mathbb{R}^{136=2\times 68}$.
- emotion representation: 2-dimensional valence-arousal ($\boldsymbol{\theta} \in \mathbb{R}^2$).
- $k_{\mathcal{X}}$, k_{Θ}: Gaussian kernels.
- Performance metrics:
  - test MSE – direct measure,
  - classification accuracy – indirect evaluation (ResNet-18 classifier; cross).

Quantitative illustration: setting

- 2 popular facial benchmarks: KDEF, RaFD.
- 68 2D landmarks: $M = 68, \mathbf{x} \in \mathbb{R}^{136=2\times 68}$.
- emotion representation: 2-dimensional valence-arousal ($\boldsymbol{\theta} \in \mathbb{R}^2$).
- $k_{\mathcal{X}}$, k_{Θ}: Gaussian kernels.
- Performance metrics:
  - test MSE – direct measure,
  - classification accuracy – indirect evaluation (ResNet-18 classifier; cross).
- Baseline: StarGAN
  - discrete emotion labels,
  - modified to handle landmarks (=: Landmark-StarGAN),
  - unstable training (as usual).

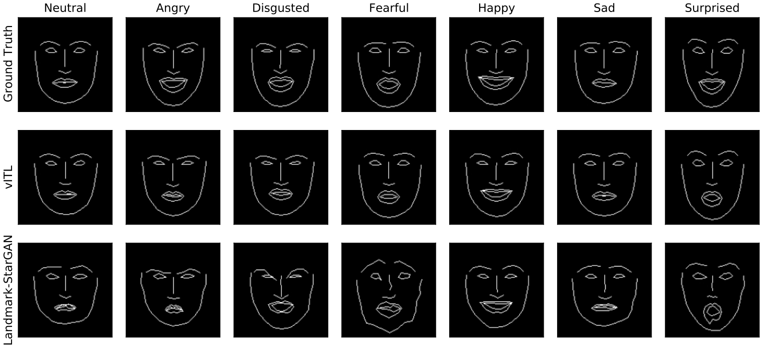
Quantitative illustration: mean $\pm$ std

Methods	MSE Error $\downarrow$		Emotion Classification Acc. $\uparrow$	
	KDEF frontal	RaFD frontal	KDEF frontal	RaFD frontal
vITL: θ_0 = neutral	0.010 ± 0.001	0.009 ± 0.004	76.12 ± 4.57	79.76 ± 7.88
vITL: θ_0 = fearful	0.010 ± 0.001	0.010 ± 0.005	76.22 ± 4.91	78.81 ± 8.36
vITL: θ_0 = angry	0.012 ± 0.002	0.010 ± 0.005	74.49 ± 2.31	78.10 ± 7.51
vITL: θ_0 = disgusted	0.012 ± 0.001	0.010 ± 0.004	74.18 ± 4.22	78.33 ± 4.12
vITL: θ_0 = happy	0.011 ± 0.001	0.010 ± 0.004	73.57 ± 2.74	80.48 ± 5.70
vITL: θ_0 = sad	0.011 ± 0.001	0.009 ± 0.004	75.82 ± 4.11	77.62 ± 5.17
vITL: θ_0 = surprised	0.010 ± 0.001	0.011 ± 0.006	74.69 ± 2.25	80.71 ± 5.99
vITL: Joint	$\mathbf{0.011} \pm 0.001$	$\mathbf{0.007} \pm 0.001$	$\mathbf{74.81} \pm 3.10$	$\mathbf{77.11} \pm 3.97$
Landmark-StarGAN	0.029 ± 0.003	0.024 ± 0.007	70.69 ± 8.46	65.88 ± 8.92

Both MSE and classification accuracy improve.

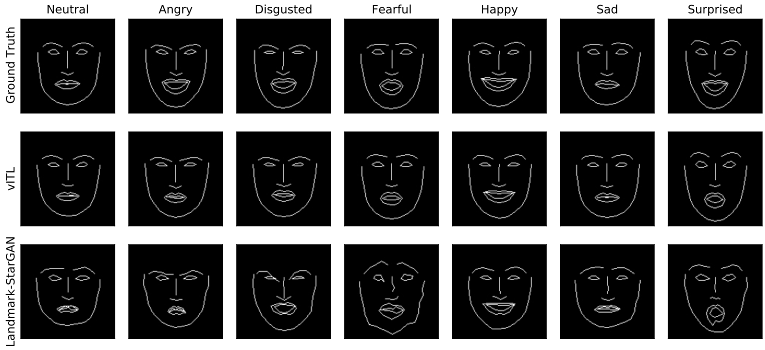
Qualitative illustration

Vs Landmark-StarGAN:

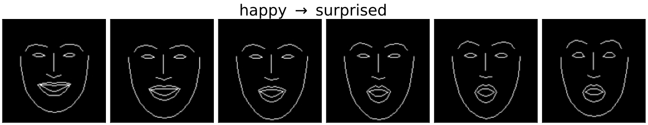


Qualitative illustration

Vs Landmark-StarGAN:



Continuous traversal by vITL ($\hat{h}$):



Summary

- We considered a new task: continuous style transfer.
- Phrasing: functional output regression.
- Model:
  - vv-RKHS framework: $\mathcal{X} \mapsto (\Theta \mapsto \mathcal{Y})$,
  - general umbrella: similarity on object/style space.
- Application: emotion transfer.

- **Acks**: 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.
- Emotion transfer by vITL (CtrlGen @ NeurIPS-2021).
- ITL @ AISTATS-2019 [Brault et al., 2019].
- Code: `https://github.com/allambert/torch_itl`.

- **Acks**: 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.
- Emotion transfer by vITL (CtrlGen @ NeurIPS-2021).
- ITL @ AISTATS-2019 [Brault et al., 2019].
- Code: `https://github.com/allambert/torch_itl`.

Thank you for the attention!

Aberman, K., Weng, Y., Lischinski, D., Cohen-Or, D., and Chen, B. (2020).
Unpaired motion style transfer from video to animation.
ACM Transactions on Graphics (TOG), 39(4):64.

Aristidou, A., Zeng, Q., Stavrakis, E., Yin, K., Cohen-Or, D., Chrysanthou, Y., and Chen, B. (2017).
Emotion control of unstructured dance movements.
In *ACM SIGGRAPH/Eurographics Symposium on Computer Animation*, pages 1–10.

Bai, L., Cui, L., Rossi, L., Xu, L., Bai, X., and Hancock, E. (2020).
Local-global nested graph kernels using nested complexity traces.
Pattern Recognition Letters, 134:87–95.

Balaei, A. T., Sutherland, K., Cistulli, P. A., and de Chazal, P. (2017).

Automatic detection of obstructive sleep apnea using facial images.
In *IEEE International Symposium on Biomedical Imaging (ISBI)*, pages 215–218.

Barricelli, B. R., Casiraghi, E., and Fogli, D. (2019).
A survey on digital twin: Definitions, characteristics, applications, and design implications.
IEEE Access, 7:167653–167671.

Berlinet, A. and Thomas-Agnan, C. (2004).
Reproducing Kernel Hilbert Spaces in Probability and Statistics.
Kluwer.

Borgwardt, K., Ghisu, E., Llinares-López, F., O'Bray, L., and Rieck, B. (2020).
Graph kernels: State-of-the-art and future challenges.
Foundations and Trends in Machine Learning, 13(5-6):531–712.

Borgwardt, K. M. and Kriegel, H.-P. (2005).
Shortest-path kernels on graphs.
In *International Conference on Data Mining (ICDM)*, pages 74–81.

Brault, R., Lambert, A., Szabó, Z., Sangnier, M., and d'Alché-Buc, F. (2019).
Infinite task learning in RKHSs.
In *International Conference on Artificial Intelligence and Statistics (AISTATS)*, pages 1294–1302.

Choi, Y., Choi, M., Kim, M., Ha, J.-W., Kim, S., and Choo, J. (2018).
StarGAN: Unified generative adversarial networks for multi-domain image-to-image translation.
In *Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 8789–8797.

Collins, M. and Duffy, N. (2001).
Convolution kernels for natural language.

In *Advances in Neural Information Processing Systems (NIPS)*, pages 625–632.

Cuturi, M. (2011).
Fast global alignment kernels.
In *International Conference on Machine Learning (ICML)*, pages 929–936.

Cuturi, M., Fukumizu, K., and Vert, J.-P. (2005).
Semigroup kernels on measures.
Journal of Machine Learning Research, 6:1169–1198.

Cuturi, M. and Vert, J.-P. (2005).
The context-tree kernel for strings.
Neural Networks, 18(8):1111–1123.

Cuturi, M., Vert, J.-P., Birkenes, O., and Matsui, T. (2007).
A kernel for time series based on global alignments.
In *International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pages 413–416.

Devries, T., Biswaranjan, K., and Taylor, G. W. (2014).
Multi-task learning of facial landmarks and expression.
In *Canadian Conference on Computer and Robot Vision*, pages 98–103.

Fu, Z., Tan, X., Peng, N., Zhao, D., and Yan, R. (2018).
Style transfer in text: Exploration and evaluation.
In *Conference on Artificial Intelligence (AAAI)*, pages 663–670.

Gärtner, T., Flach, P., Kowalczyk, A., and Smola, A. (2002).
Multi-instance kernels.
In *International Conference on Machine Learning (ICML)*, pages 179–186.

Gärtner, T., Flach, P., and Wrobel, S. (2003).
On graph kernels: Hardness results and efficient alternatives.
Learning Theory and Kernel Machines, pages 129–143.

Grinstein, E., Duong, N. Q., Ozerov, A., and Pérez, P. (2018).
Audio style transfer.

In *International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 586–590.

Guevara, J., Hirata, R., and Canu, S. (2017).
Cross product kernels for fuzzy set similarity.
In *International Conference on Fuzzy Systems (FUZZ-IEEE)*, pages 1–6.

Haussler, D. (1999).
Convolution kernels on discrete structures.
Technical report, University of California at Santa Cruz.
(http://cbse.soe.ucsc.edu/sites/default/files/convolutions.pdf).

Hein, M. and Bousquet, O. (2005).
Hilbertian metrics and positive definite kernels on probability measures.
In *International Conference on Artificial Intelligence and Statistics (AISTATS)*, pages 136–143.

Irimia, A.-S., Chan, J. C., Mistry, K., Wei, W., and Ho, E. S. (2019).
Emotion transfer for hand animation.
In *Motion, Interaction and Games*, pages 1–2.

Jaakkola, T. S. and Haussler, D. (1999).
Exploiting generative models in discriminative classifiers.
In *Advances in Neural Information Processing Systems (NIPS)*, pages 487–493.

Jebara, T., Kondor, R., and Howard, A. (2004).
Probability product kernels.
Journal of Machine Learning Research, 5:819–844.

Jiao, Y. and Vert, J.-P. (2016).
The Kendall and Mallows kernels for permutations.
In *International Conference on Machine Learning (ICML)*, volume 37, pages 2982–2990.

Kashima, H. and Koyanagi, T. (2002).

Kernels for semi-structured data.
In *International Conference on Machine Learning (ICML)*, pages 291–298.

Kashima, H., Tsuda, K., and Inokuchi, A. (2003).
Marginalized kernels between labeled graphs.
In *International Conference on Machine Learning (ICML)*, pages 321–328.

Király, F. J. and Oberhauser, H. (2019).
Kernels for sequentially ordered data.
Journal of Machine Learning Research, 20:1–45.

Kondor, R. and Pan, H. (2016).
The multiscale Laplacian graph kernel.
In *Advances in Neural Information Processing Systems (NIPS)*, pages 2982–2990.

Kondor, R. I. and Lafferty, J. (2002).
Diffusion kernels on graphs and other discrete input.

In *International Conference on Machine Learning (ICML)*, pages 315–322.

Kuang, R., Ie, E., Wang, K., Wang, K., Siddiqi, M., Freund, Y., and Leslie, C. (2004).
Profile-based string kernels for remote homology detection and motif extraction.
Journal of Bioinformatics and Computational Biology, 13(4):527–550.

Leslie, C., Eskin, E., and Noble, W. S. (2002).
The spectrum kernel: A string kernel for SVM protein classification.
Biocomputing, pages 564–575.

Leslie, C. and Kuang, R. (2004).
Fast string kernels using inexact matching for protein sequences.
Journal of Machine Learning Research, 5:1435–1455.

Lim, K. Y. H., Zheng, P., and Che, C.-H. (2020).

A state-of-the-art survey of digital twin: techniques, engineering product lifecycle management and business innovation perspectives.
Journal of Intelligent Manufacturing, 31:1313–1337.

Lodhi, H., Saunders, C., Shawe-Taylor, J., Cristianini, N., and Watkins, C. (2002).
Text classification using string kernels.
Journal of Machine Learning Research, 2:419–444.

Mane, S. and Shah, G. (2019).
Facial recognition, expression recognition, and gender identification.
In *Data Management, Analytics and Innovation*, pages 275–290. Springer.

Puy, G. and Pérez, P. (2019).
A flexible convolutional solver for fast style transfers.
In *Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 8963–8972.

Rüping, S. (2001).
SVM kernels for time series analysis.
Technical report, University of Dortmund.
(http://www.stefan-rueping.de/publications/rueping-2001-a.pdf).

Russell, J. A. (1980).
A circumplex model of affect.
Journal of Personality and Social Psychology, 39(6):1161–1178.

Saigo, H., Vert, J.-P., Ueda, N., and Akutsu, T. (2004).
Protein homology detection using string alignment kernels.
Bioinformatics, 20(11):1682–1689.

Seeger, M. (2002).
Covariance kernels from Bayesian generative models.
In *Advances in Neural Information Processing Systems (NIPS)*, pages 905–912.

Shervashidze, N., Vishwanathan, S. V. N., Petri, T., Mehlhorn, K., and Borgwardt, K. M. (2009).
Efficient graphlet kernels for large graph comparison.
In *International Conference on Artificial Intelligence and Statistics (AISTATS)*, pages 488–495.

Smola, A., Gretton, A., Song, L., and Schölkopf, B. (2007).
A Hilbert space embedding for distributions.
In *Algorithmic Learning Theory (ALT)*, pages 13–31.

Sriperumbudur, B., Gretton, A., Fukumizu, K., Schölkopf, B., and Lanckriet, G. (2010).
Hilbert space embeddings and metrics on probability measures.
Journal of Machine Learning Research, 11:1517–1561.

Tao, F., Zhang, H., Liu, A., and Nee, A. Y. C. (2019).
Digital twin in industry: State-of-the-art.
IEEE Transactions on Industrial Informatics, 15(4):2405 – 2415.

Tsuda, K., Kin, T., and Asai, K. (2002).
Marginalized kernels for biological sequences.
Bioinformatics, 18:268–275.

Ulyanov, D., Lebedev, V., Vedaldi, A., and Lempitsky, V. (2016).
Texture networks: Feed-forward synthesis of textures and stylized images.
In *International Conference on Machine Learning (ICML)*, pages 1349–1357.

Vemulapalli, R. and Agarwala, A. (2019).
A compact embedding for facial expression similarity.
In *Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5683–5692.

Vishwanathan, S. N., Schraudolph, N., Kondor, R., and Borgwardt, K. (2010).
Graph kernels.
Journal of Machine Learning Research, 11:1201–1242.

Watkins, C. (1999).
Dynamic alignment kernels.
In *Advances in Neural Information Processing Systems (NIPS)*, pages 39–50.

Yao, X., Puy, G., Newson, A., Gousseau, Y., and Hellier, P. (2020).
High resolution face age editing.
In *International Conference on Pattern Recognition (ICPR)*.