A Robust Person Shape Representation via Grassmann Channel Pooling

Matsukawa, Tetsu
Kyushu University

Suzuki, Einoshin
Kyushu University

A Robust Peron Shape Representation via Grassmann Channel Pooling

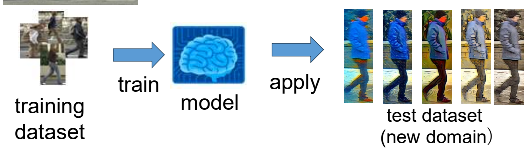
Tetsu Matsukawa and Einoshin Suzuki, Kyushu University, Japan

Motivation

Person orientation estimation

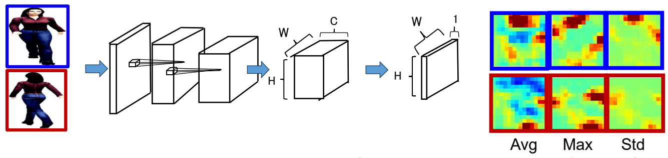
Domain generalization

Improve system's robustness in different conditions from training datasets



Channel pooling

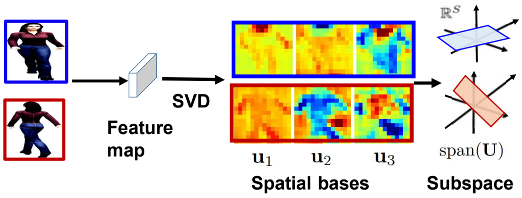
Extract shape information by summarizing a feature map along the channel dimensions



Naïve channel pooling methods often results in loss of significant discriminative information, as they focus on only local information

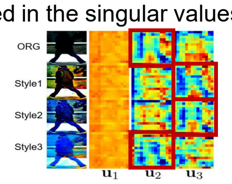
Method

Grassmann Channel Pooling (GCP)



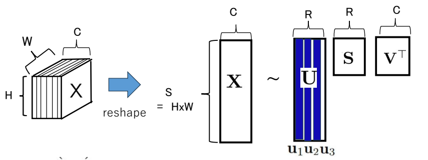
Summarize each feature map as linear subspace of spatial bases

- Each spatial base represents positions which shares globally similar property
- Activation magnitude is decoupled in the singular values
- Subspace $\mathrm{span}(\mathbf{U})$ is unique regardless the order of spatial bases
  i.e., span($[\mathbf{u}_1\mathbf{u}_2\mathbf{u}_3]$) = span($[\mathbf{u}_1\mathbf{u}_3\mathbf{u}_2]$)



Algorithm 1 Grassmann Channel Pooling (GCP)

Input: Feature tensor X, rank R



Projection metric on Grassmann manifold [Edelman SAIM'98]

$$\Phi(\mathbf{U}) = \mathbf{U}\mathbf{U}^\top$$

Output: A vectorized representation $\mathbf{z} = \mathrm{vec}(\Phi(\mathbf{U}))$

$\Phi(\mathbf{U})$ contains all co-occurrence statistics of S x S positional combinations ➡ discriminative for human pose

Robustness

- **Channel permutation invariance**
  Robust to case when a specific spatial pattern within a feature channel activates on different channels

- **Channel magnitude invariance**
  Robust to the magnitude change

Connection with Bilinear Pooling

Bilinear Spatial Pooling (BSP) [Lin CVPR'15]

$$\frac{1}{S}\mathbf{X}^\top\mathbf{X} = \mathbf{V}^{*\top}\left(\frac{1}{S}\mathbf{S}^{*2}\right)\mathbf{V}^*$$

Sensitive to image style change, as it represents image style [Gatys'16]

Grassmann Spatial Pooling (GSP) [Wei ECCV'18]

$$\mathbf{V}^\top\mathbf{V}$$

Robust against style change as it does not include $\mathbf{S}^*$

Select leading bases and neglect singular values

Bilinear Channel Pooling(BCP)

$$\frac{1}{C}\mathbf{X}\mathbf{X}^\top = \mathbf{U}^*\left(\frac{1}{C}\mathbf{S}^{*2}\right)\mathbf{U}^{*\top}$$

Robust against style change as it does not include $\mathbf{V}^*$

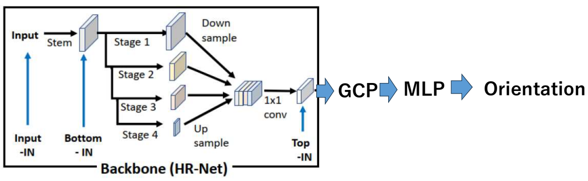
GCP

$$\mathbf{U}\mathbf{U}^\top$$

Most robust against style change as it does not include both $\mathbf{S}^*$ and $\mathbf{V}^*$

Experiments

Architecture



- Backbone: High-Resolution Net-v2 [Wang PAMI '21]
- Instance Normalization (IN) adds robustness
- Loss: Von Mises loss [Beyer GCPR'15]

Dataset

Person X TUD

ORG

Styled test dataset by a style transfer model STTR [Wang ACCV '22]

Performance comparison

	Person X					TUD				
	ORG		Styled			ORG		Styled		
Method	$Acc_{10°}$ ↑	MAE ↓	$Acc_{10°}$ ↑	MAE ↓	SIM ↑	$Acc_{10°}$ ↑	MAE ↓	$Acc_{10°}$ ↑	MAE ↓	SIM ↑
Spatial Pooling										
Avg	96.1	3.4	75.5	7.7	98.1	28.6	31.5	15.7	51.6	59.7
BSP	95.8	3.4	84.5	6.3	98.7	31.9	26.0	25.7	33.1	81.2
GSP	95.9	3.3	86.0	5.9	99.0	38.5	24.6	29.8	32.2	82.8
Channel Pooling										
Avg	94.0	4.0	82.2	6.8	98.6	34.3	26.6	27.3	36.8	75.2
Std	95.0	3.7	83.4	6.5	98.9	29.6	28.0	25.6	33.3	77.4
Max	95.2	3.6	82.7	6.6	98.8	31.4	25.9	26.4	30.3	85.7
BCP	95.2	3.5	88.2	5.4	99.3	35.5	24.2	27.6	32.1	79.3
GCP	95.6	3.4	87.2	5.6	99.2	35.8	21.6	33.1	25.5	87.6

Parameter analysis on TUD