Colors See Colors Ignore: Clothes Changing ReID with Color Disentanglement

https://ucf-crcv.github.io/ReID/CSCI

CRCV Lab, UNIVERSITY OF CENTRAL FLORIDA

Priyank Pathak

priyank@ucf.edu

Yogesh S Rawat

yogesh@ucf.edu

Problem Statement:
Clothes Changing ReID

- Identify a person irrespective of change of clothes

- Ignore appearance bias (clothes / environment)

- Real world, **people change clothes** :)

Current Research

- Clothes as "integer" labels
    - E.g. person is wearing cloth id 0
    - Predict classes (Almost all works)
    - Non-descriptive
    - Constant label across video

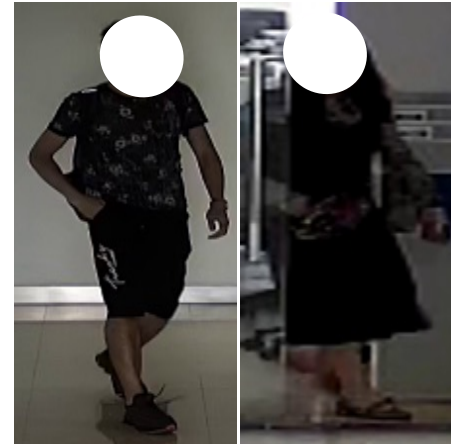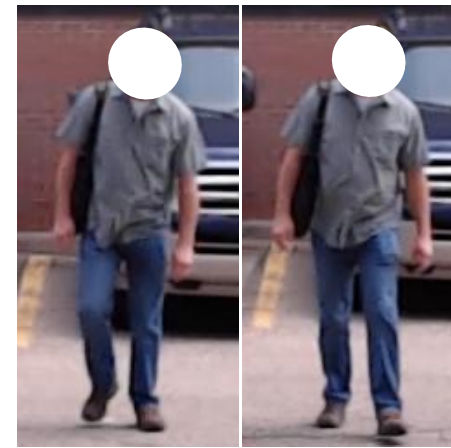LLM based fine-grained
Computationally very heavy
What if person changes appearance in video
E.g. person removes "black hat"
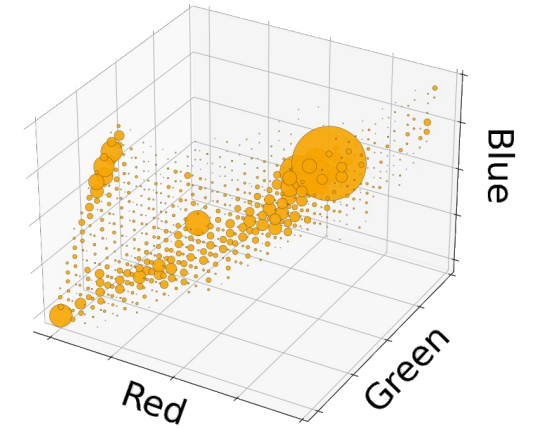Non scalable

Solution: Pixel Colors

- Colors cluster **similar clothing**

- Extract colors "efficiently"
  - **RGB-only**
  - **Free of cost**
  - **Annotation free**

- Colors are **contextual**
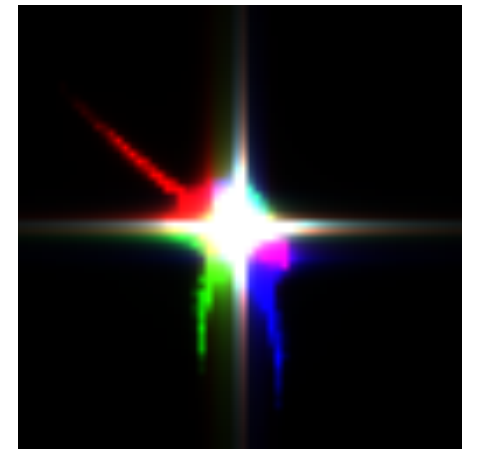
- Colors are **adaptive**

Color Representation

Pixel Binning

- "Pixel Binning" : Count # of pixels in each channel
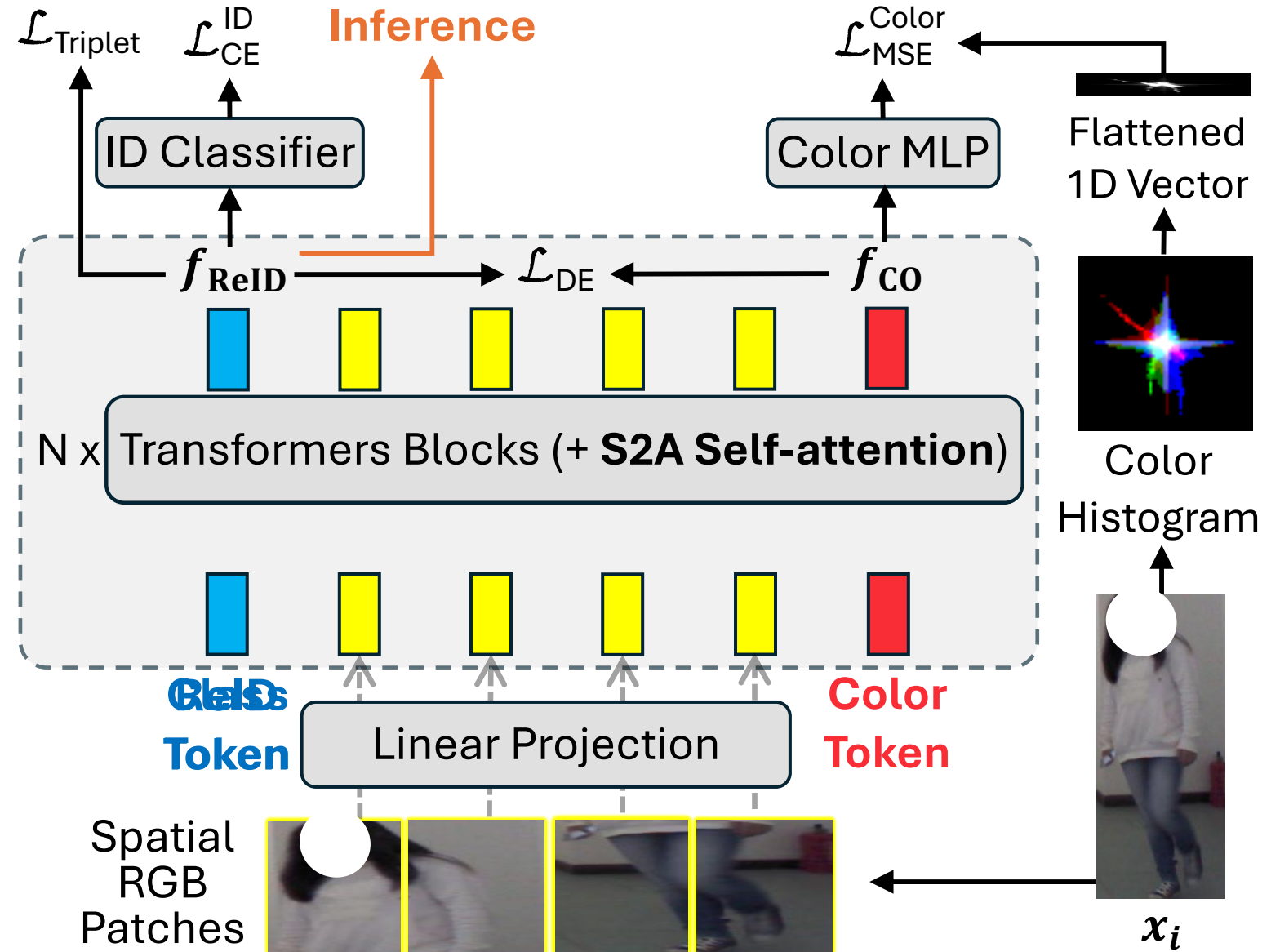
- "RGB – uv": 2D projection of colors (style GAN)

RGB-uv

Color See, Color Ignore (CSCI)

- Transformers

- Efficient

- One Model (1 branch) to ReID and disentangle

$\mathcal{L}_{\text{Triplet}}$ $\mathcal{L}_{\text{CE}}^{\text{ID}}$ **Inference**

ID Classifier

$\mathcal{L}_{\text{MSE}}^{\text{Color}}$

Color MLP

Flattened 1D Vector

f_{ReID} $\mathcal{L}_{\text{DE}}$ f_{CO}

N x Transformers Blocks (+ **S2A Self-attention**)

Color Histogram

Class Token ReID

Color Token

Linear Projection

Spatial RGB Patches

x_i

What's S2A self-attention?

- Traditional Self-attention

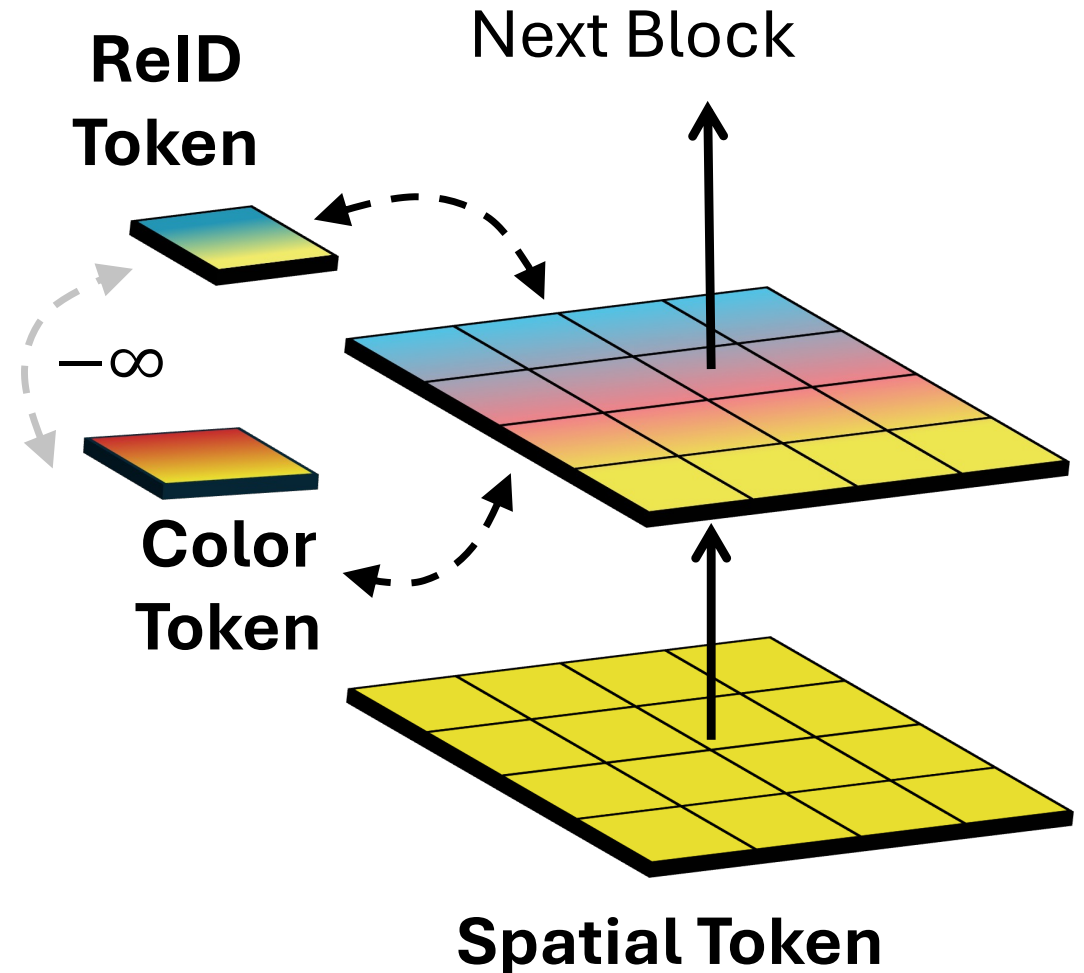- 100% information leak between ReID (biometrics) & Color tokens (appearance)
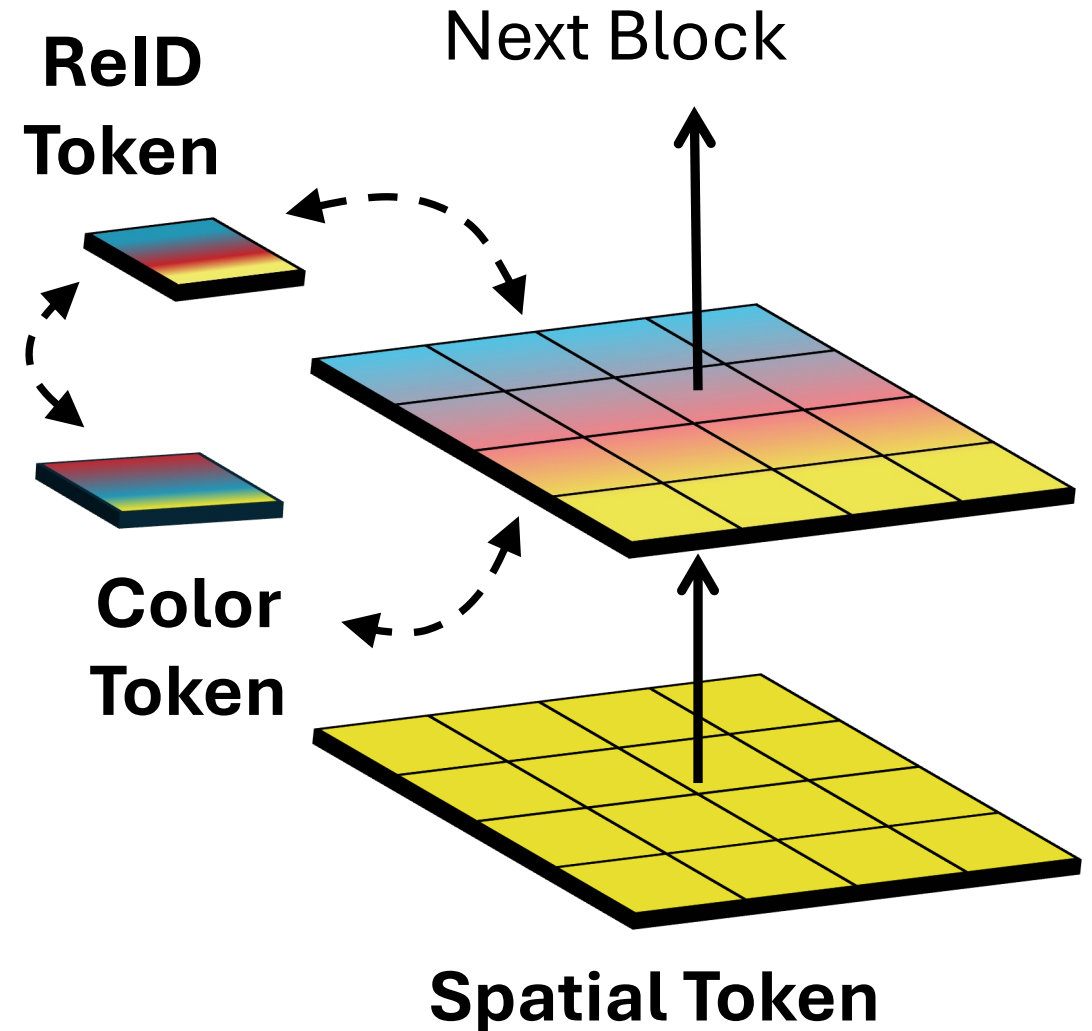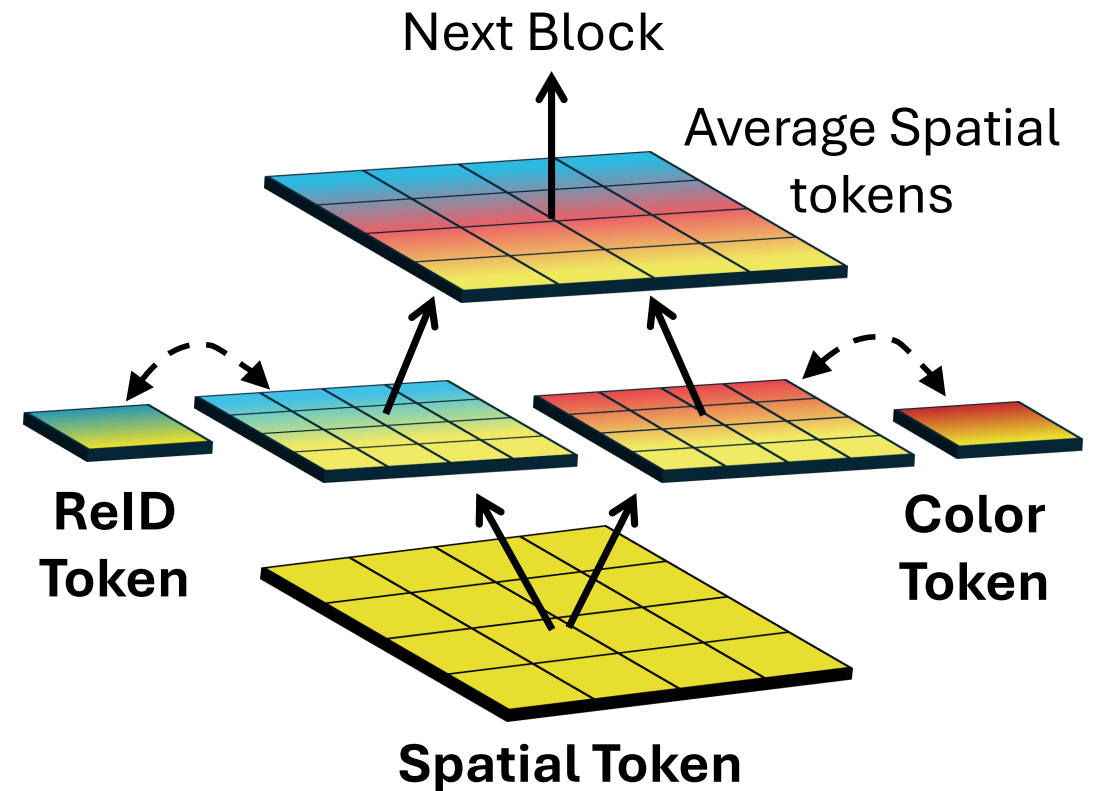
What's S2A self-attention?

- Masked Self-attention

- No direct leak

- ReID token influences weight Color token
  - Biometric affecting weight of appearance bias
  - & vice versa

Ours S2A self-attention?

- Two self-attention steps

- No direct leak

- **DO NOT** affect the weight

- Alternative: 2 transformer branches (infeasible)

Results

- Color tokens Helps

- Masked (no direct leak) helps

- S2A outperforms Masked
  - 4.4% gain in FLOP

Self-Attention	GFlop	LTCC	PRCC
No Color Token	81.18	44.9	61.6
Traditional	81.52	46.8	63.5
Masked	81.52	46.7	65.3
S2A (our)	84.74	47.8	66.2

Results

- Works well with different backbones

Backbone	Self-Attention	LTCC
ViT-B	TCiP	35.2
	+ Clothes	36.5
	+ CSCI (our)	38.8
EVA-02	Baseline	44.9
	+ Clothes	46.3
	+ CSCI (our)	47.8

Backbone	Self-Attention	LTCC
ViT-B	TransReID	31.1
	+ CSCI (our)	36.0
	PAT	31.1
	+ CSCI (our)	32.7
ViT-S	TMGF	32.9
	+ CSCI (our)	39.5

- Outperform "clothes integer labels"

Overall Image Results

	Method	Venue	Additional Attributes		LTCC CC		LTCC General		PRCC CC		PRCC SC	
			Int.	External	R-1↑	mAP↑	R-1↑	mAP↑	R-1↑	mAP↑	R-1↑	mAP↑
ResNet	CAL [16]	CVPR'22	CL		40.1	18.0	74.2	40.8	55.2	55.8	**100**	99.8
ResNet	CCFA [18]	CVPR'23	CL		45.3	22.1	75.8	42.5	61.2	58.4	99.6	98.7
ResNet	FIRe^2 [41]	TIFS'24	-	-	44.6	19.1	75.9	39.9	65.0	63.1	**100**	99.5
Multi-ResNets	3DInvar. [28]	ICCV'23	CL	Po+BS	40.9	18.9	-	-	56.5	57.2	-	-
Multi-ResNets	AIM [44]	CVPR'23	CL		40.6	19.1	76.3	41.1	57.9	58.3	**100**	**99.9**
Multi-ResNets	DCR. [10]	TCSVT'23	CL	BP+Co.	41.1	20.4	76.1	42.3	57.2	57.4	**100**	99.7
Multi-ResNets	CVSL [31]	WACV'24	CL	Po	44.5	21.3	76.4	41.9	57.5	56.9	97.5	99.1
Multi-ResNets	CCPG [32]	CVPR'24	CL	Po	46.2	22.9	77.2	42.9	61.8	58.3	**100**	99.6
Transformer	CLIP3D [29]	CVPR'24	-	BS+Text	42.1	21.7	-	-	60.6	59.3	-	-
Transformer	IRM [20]†	CVPR'24	CL	Text+BP	46.7	-	66.7	-	-	-	-	-
Transformer	EVA-02		-	-	44.9	23.1	80.3	45.9	61.6	59.0	**100**	**99.9**
Transformer	CSCI - Pix. Bin (our)		-	-	46.7+1.8	23.6+0.5	80.8+0.5	45.9	**66.6**+5.0	60.7+1.7	**100**	**99.9**
Transformer	CSCI - RGB-uv (our)		-	-	**47.8**+2.9	**24.4**+1.3	82.6+2.3	48.0+2.1	66.2+4.6	**61.3**+2.3	**100**	**99.9**