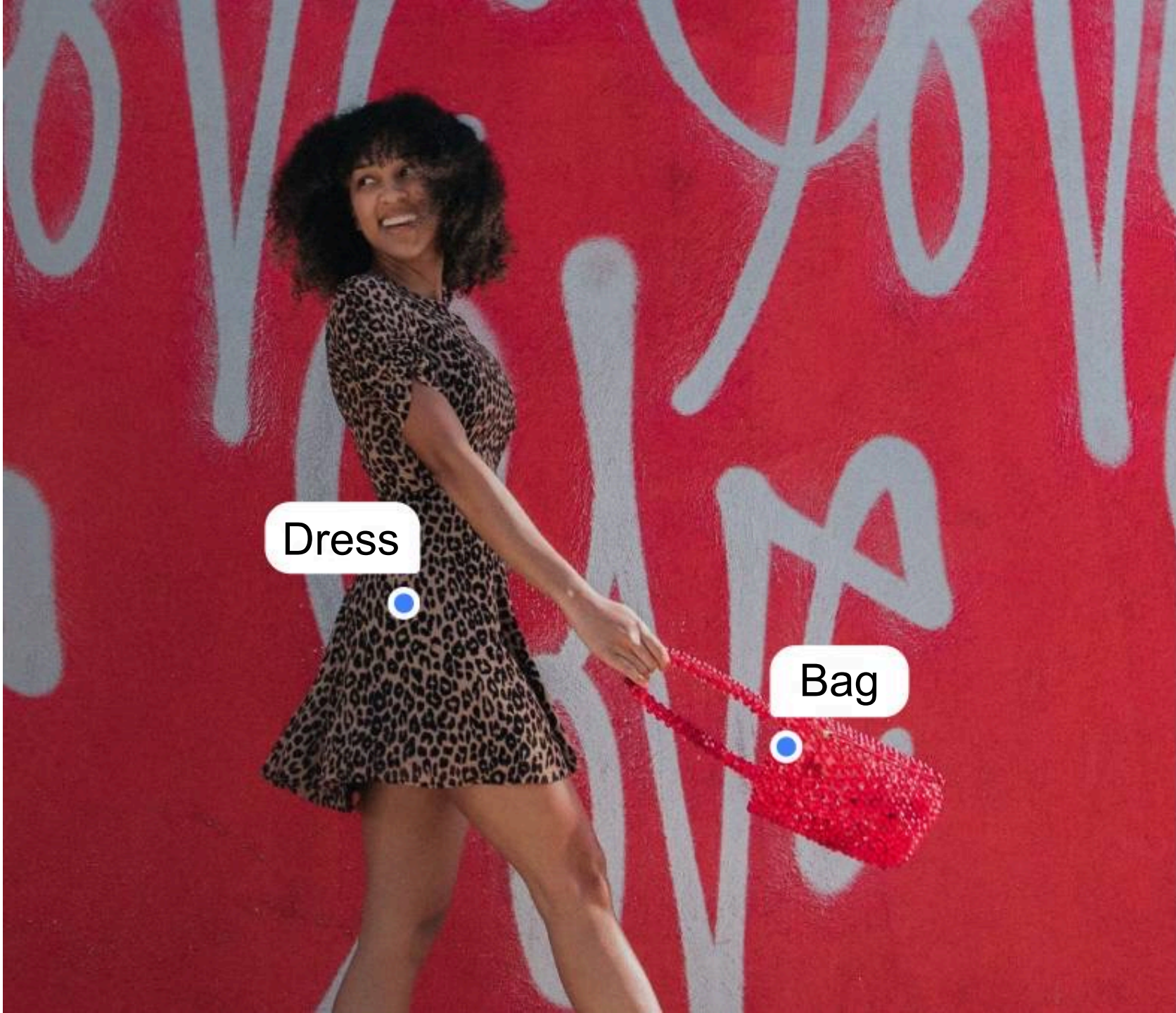


# MiVOLO

Multi-input Transformer for Age and Gender Estimation

Maksim Kuprashevich & Irina Tolstykh





Female, 28

Dress

Bag



Leopard dress  
[Buy](#)



Red handbag  
[Buy](#)



# Problem description



Age ?  
Gender ?

## Open datasets

- Regression
  - **IMDB-clean**: 183,886 training, 45,971 validation and 56,086 test images
  - **UTKFace**: 13,144 training and 3,287 validation images
  - **AgeDB**: 16,488 images
- Classification
  - **FairFace**: 86,744 training and 10,954 validation images
  - **Adience**: 26,580 images



## Age metrics

$$\frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

Mean Absolute Error (**MAE**)

$$CS_l = \frac{N_l}{N} \times 100\%$$

**N** - total number of testing examples


**N<sub>l</sub>** - count of examples for which the absolute error between the estimated age and the true age does not exceed 1 years

Cumulative Score (**CS@1**)

# LAGENDA Dataset

## Crowd Source

Overlap 10



Unable to determine gender, age, or if there are multiple people in the photo [!].

Please estimate the approximate age of the person in the photo (+/- 5 years).

Please indicate the gender of the person in the photo.

Skip

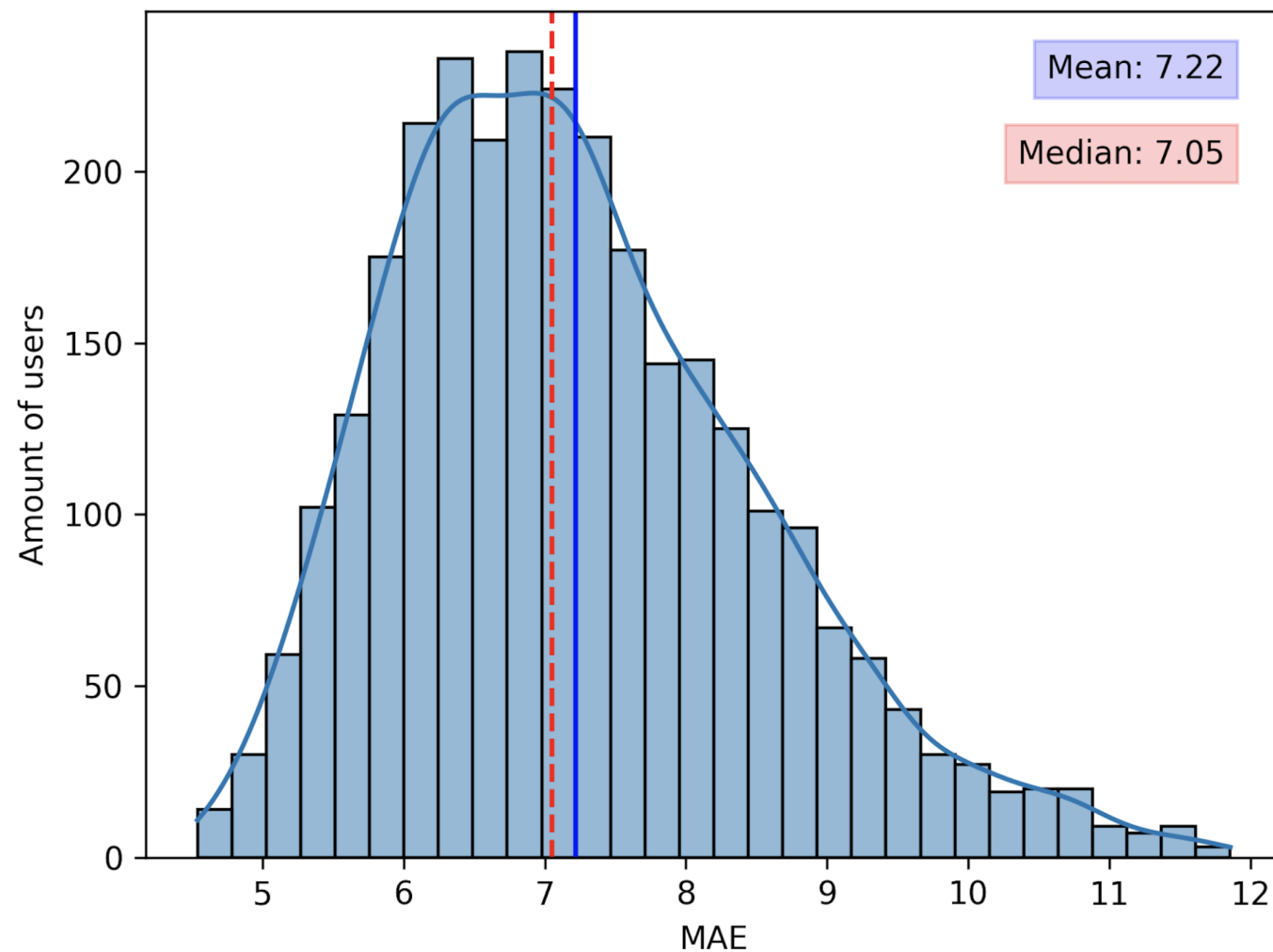
Open Images Dataset

Production Data

Additional sources

IMDB-clean  
as control tasks

# Annotators assessment quality



Histogram of MAE across users, measured on control tasks ( $n \geq 20$ ).

# Crowdsource data aggregation

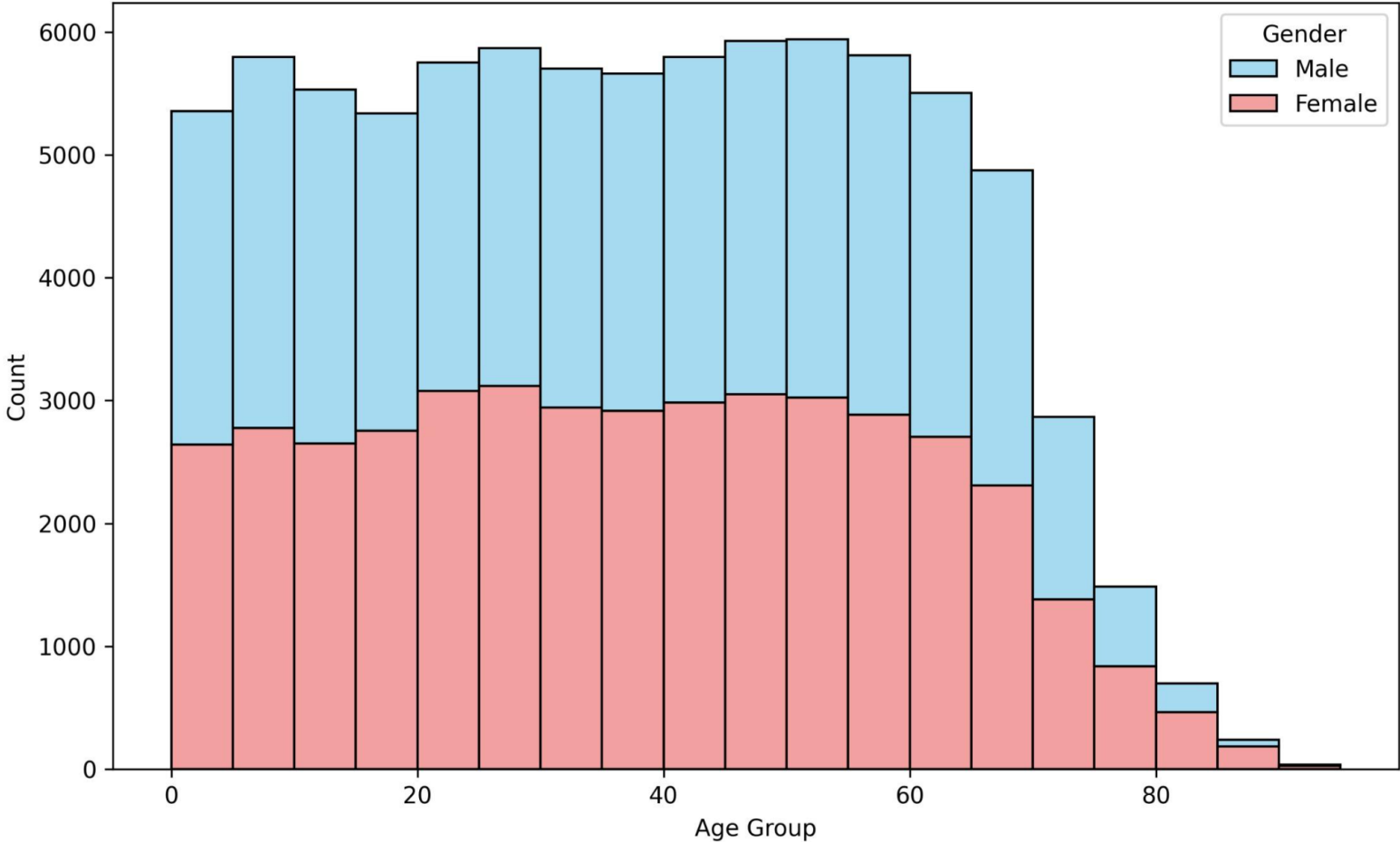
Method	MAE	CS@5, %
Mean	4.77	62.43
Median	4.75	65.44
Interquartile Mean	4.74	63.80
Mode	5.70	59.28
Maximum Likelihood	4.81	65.86
Winsorized mean (6)	4.73	63.44
Truncated mean (0.3)	4.75	63.62
<b>Weighted mean</b>	<b>3.47</b>	<b>74.31</b>

Table 1: Different statistic methods to aggregate  $N$  votes into one age prediction.

$$A(v) = \frac{\sum_{i=1}^N v_i * e^{(MAE(u_i))^{-1}}}{\sum_{i=1}^N e^{(MAE(u_i))^{-1}}}$$

where  $A$  is final age prediction for the  $v$  vector of user votes,  $N$  is size of  $v$ , amount of users who annotated this sample and  $MAE(u_i)$  denotes the individual MAE all control tasks for the  $i$ -th user  $u$ .

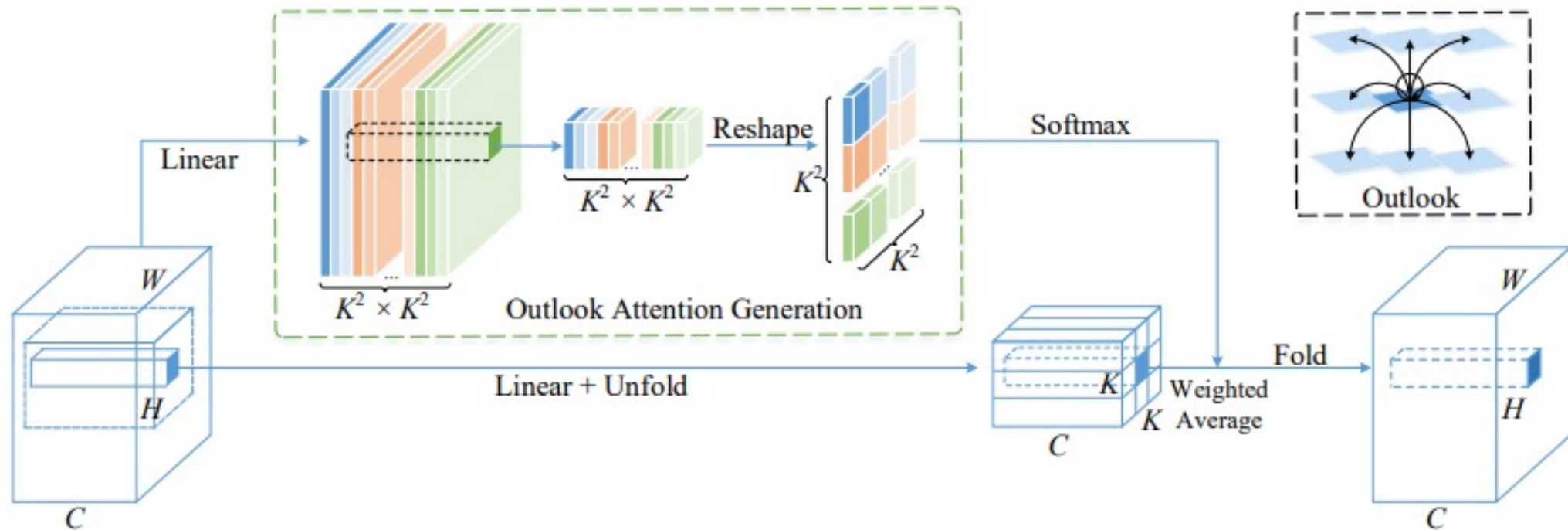
# LAGENDA: data distribution



Open LAGENDA  
Benchmark:  
67,159 images with  
84,192 persons  
  
~500k train images

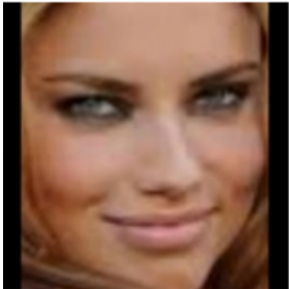
<https://wildchlamydia.github.io/lagenda/>

# VOLO: Vision Outlooker for Visual Recognition

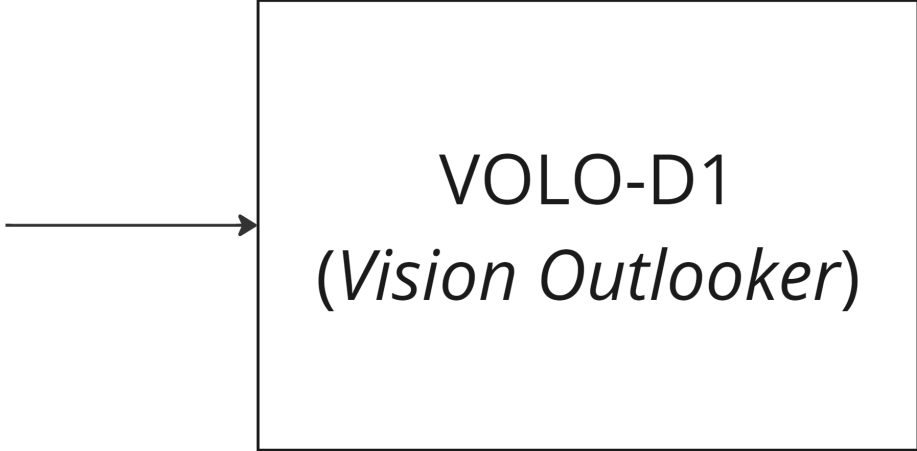


<https://arxiv.org/abs/2106.13112>

# Baseline



Face Crop  
3 x 224 x 224



Age  
*Regression*

Gender  
*Classification*

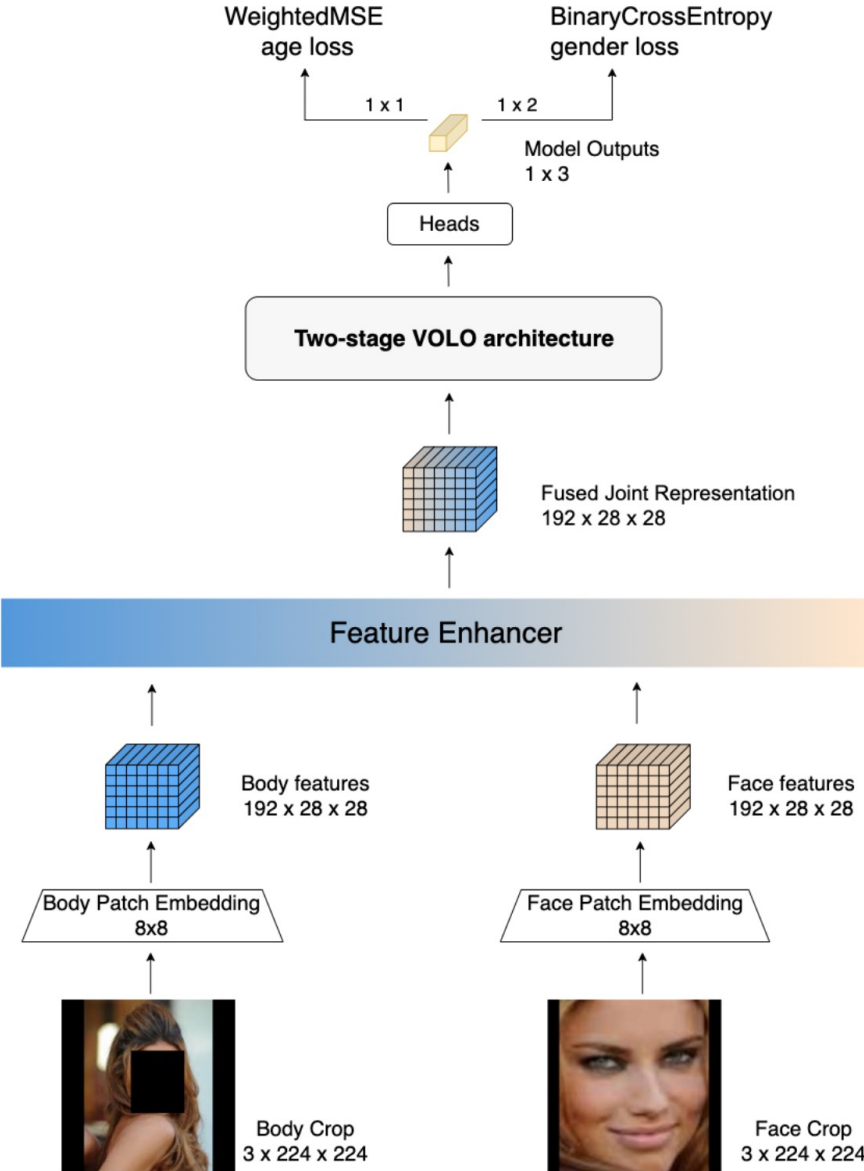
# VOLO is the SOTA already

Model	Output	Train Dataset	Test Dataset	Age MAE	Age CS@5	Gender Acc
FP-Age [27]	age	IMDB-clean	IMDB-clean	4.68	63.78	-
VOLO-D1 face†	age	IMDB-clean	IMDB-clean	4.29	67.71	-
			UTKFace	5.28	56.79	-
			Lagenda test	5.46	57.90	-
VOLO-D1 face†	age & gender	IMDB-clean	<b>IMDB-clean</b>	<b>4.22</b>	<b>68.68</b>	<b>99.38</b>
			UTKFace	5.15	56.79	97.54
			Lagenda test	5.33	59.17	90.86
CORAL [7]	age	UTKFace	UTKFace	5.39	-	-
Randomized Bins [5]	age	UTKFace	UTKFace	4.55	-	-
MWR [35]	age	UTKFace	UTKFace	4.37	-	-
VOLO-D1 face†	age	UTKFace	IMDB-clean	8.59	37.96	-
			UTKFace	4.23	69.72	-
			Lagenda test	11.16	30.51	-
VOLO-D1 face†	age & gender	UTKFace	IMDB-clean	8.06	41.72	97.05
			<b>UTKFace</b>	<b>4.23</b>	<b>69.78</b>	<b>97.69</b>
			Lagenda test	11.37	30.20	83.27
VOLO-D1 face	age	Lagenda train	IMDB-clean	4.13	69.33	-
			UTKFace	3.90	72.25	-
			Lagenda test	4.19	69.36	-
VOLO-D1 face	age & gender	Lagenda train	IMDB-clean	<b>4.10</b>	<b>69.71</b>	<b>99.57</b>
			UTKFace	<b>3.82</b>	<b>72.64</b>	<b>98.87</b>
			<b>Lagenda test</b>	<b>4.11</b>	<b>70.11</b>	<b>96.89</b>

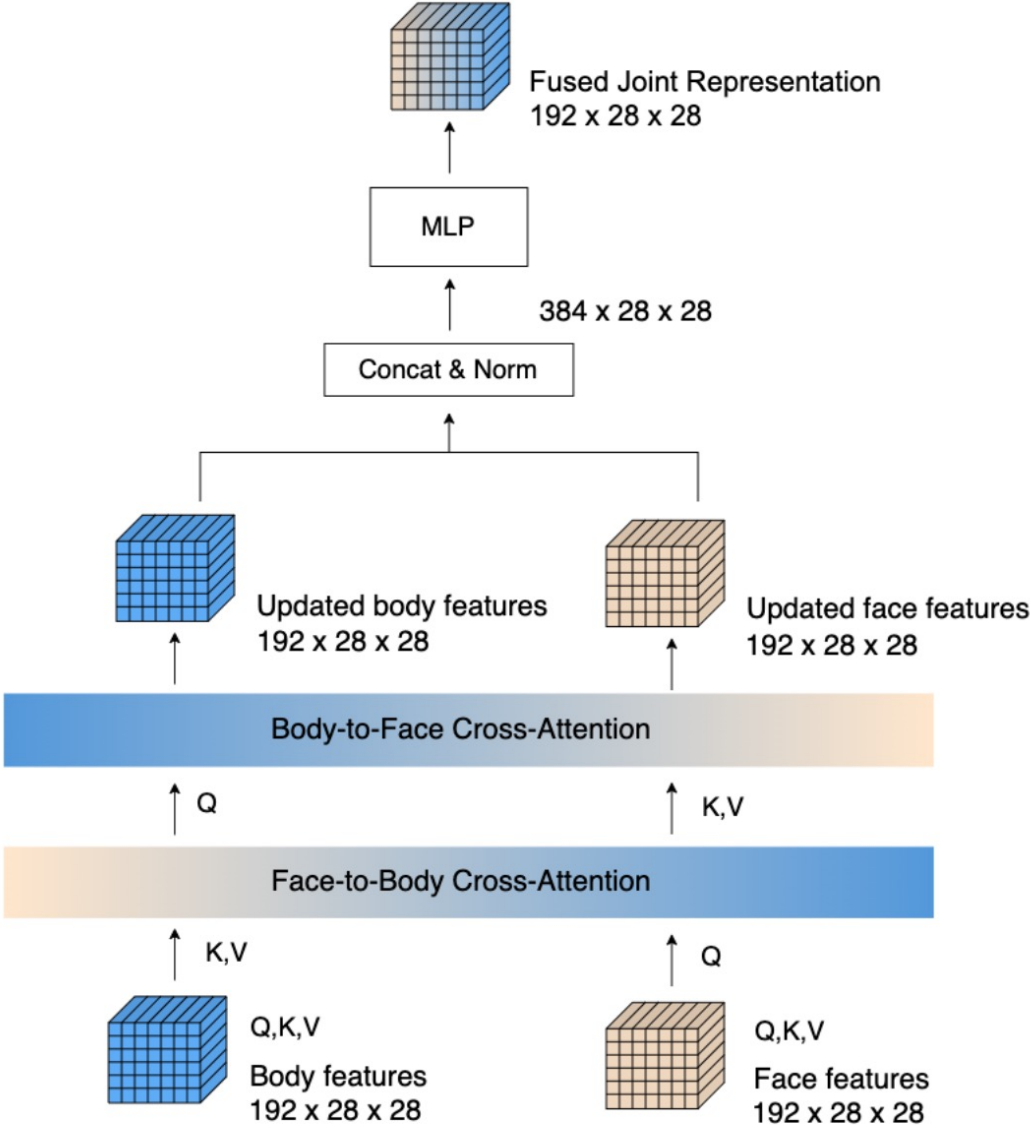
Table 2: Comparison of accuracy of VOLO-D1 models and previous SOTA results. **Bold** indicates the best model, trained and evaluated on the same datasets. **Bold** indicates the best model with additional train data. † marks the models that we release to the public domain.



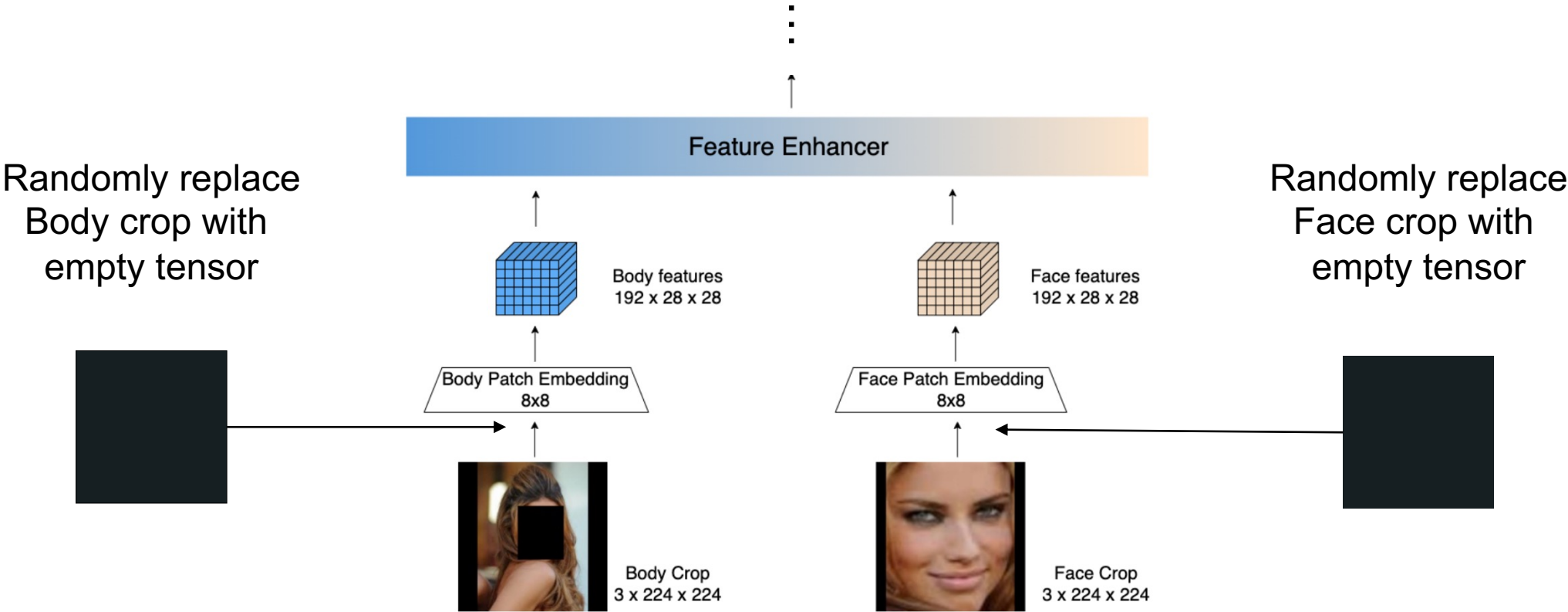
# MiVOLO



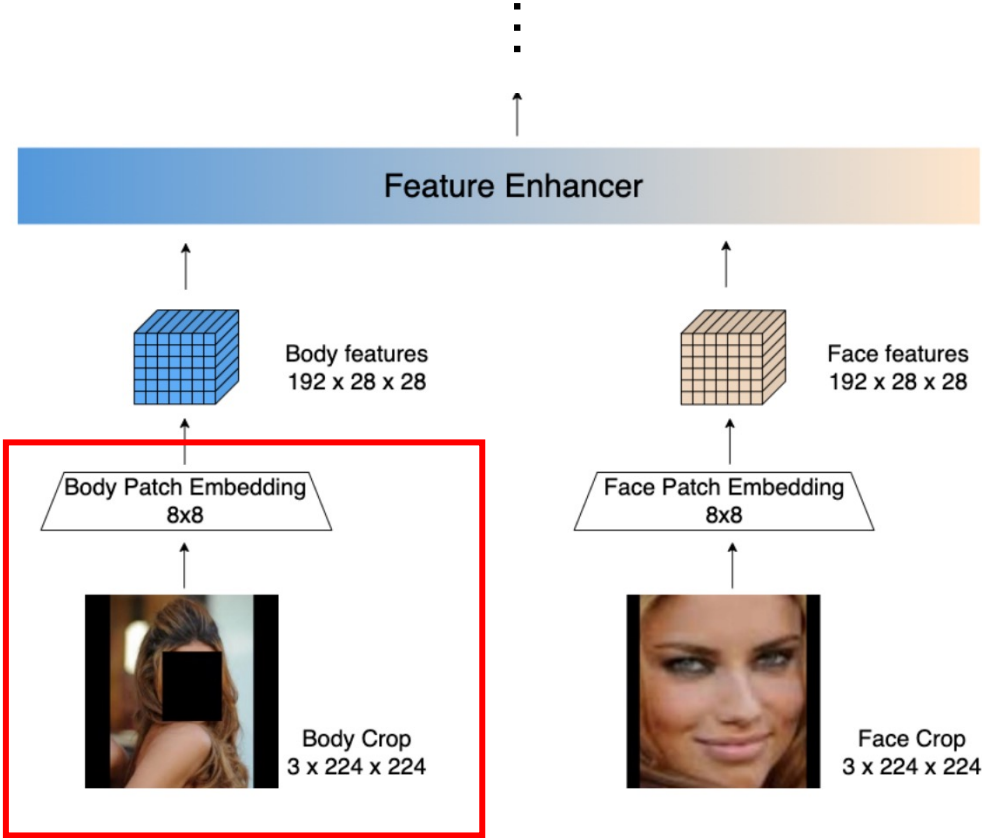
# Feature Enhancer



# Training Process: Input Dropout



# Training Process: Face bbox crop



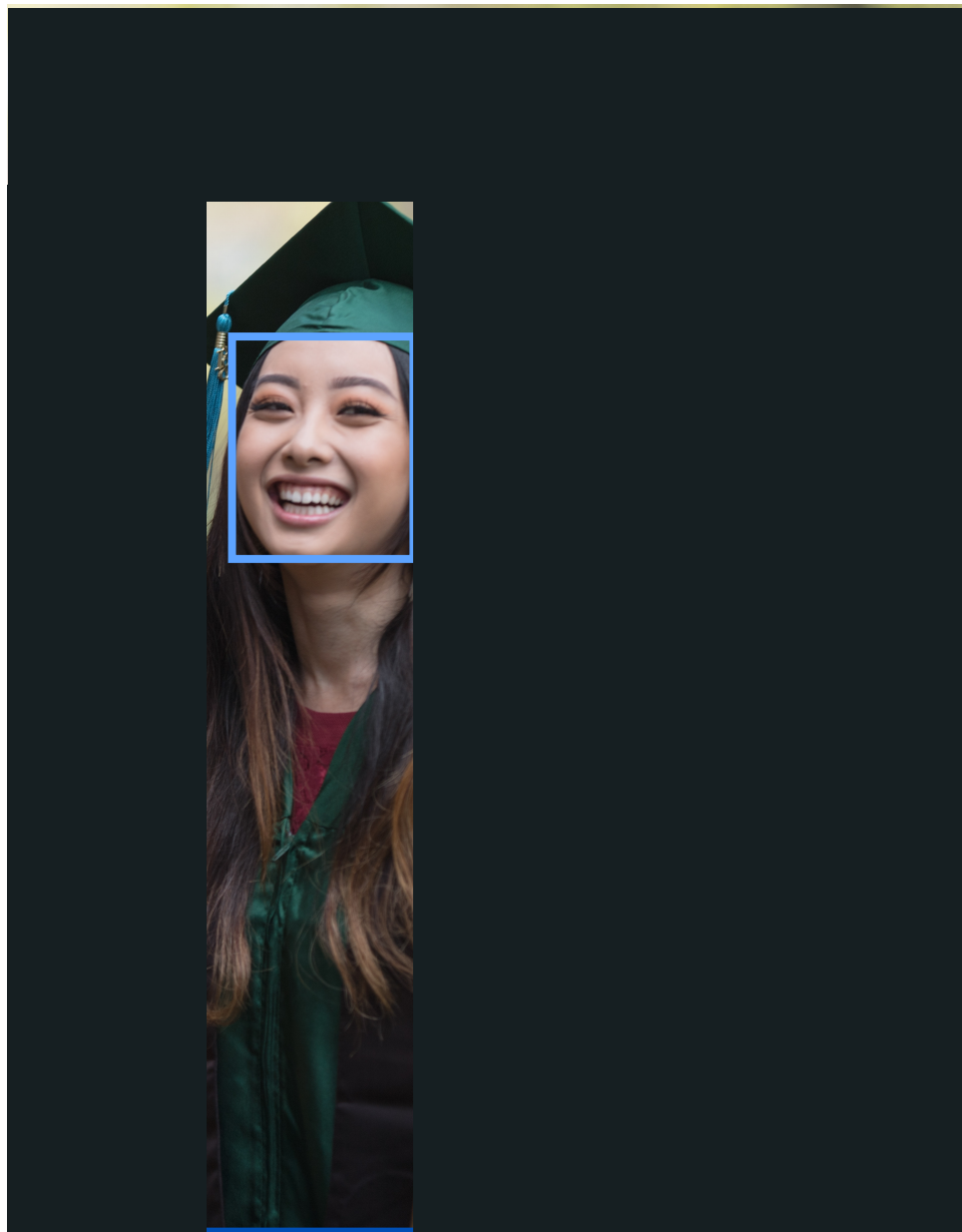
# Person trimming



# Person trimming



# Person trimming



# VOLO and MiVOLO comparison

			Tested with								
			Face			Body			Face&Body		
Model	Train Set	Test Set	MAE	CS@5	Gender Acc	MAE	CS@5	Gender Acc	MAE	CS@5	Gender Acc
VOLO-D1	Lagenda	IMDB	4.10	69.71	<b>99.57</b>	-	-	-	-	-	-
		UTKFace	3.82	72.64	<b>98.87</b>	-	-	-	-	-	-
		Lagenda	4.11	70.11	96.89	-	-	-	-	-	-
MiVOLO-D1†	IMDB	IMDB	4.35	67.18	99.39	6.87	46.32	96.48	4.24	68.32	99.46
		UTKFace	5.12	59.10	97.66	6.36	47.74	95.57	5.10	97.72	59.46
		Lagenda	5.40	58.67	91.06	10.52	31.70	87.71	5.33	59.20	91.91
MiVOLO-D1	Lagenda	IMDB	4.15	69.20	99.52	6.66	47.53	96.74	<b>4.09</b>	<b>69.72</b>	99.55
		UTKFace	3.86	72.06	98.81	4.62	63.81	98.69	<b>3.70</b>	<b>74.16</b>	98.84
		Lagenda	4.09	70.23	96.72	7.41	49.64	93.57	<b>3.99</b>	<b>71.27</b>	<b>97.36</b>

Table 3: Comparison of multi-input MiVOLO-D1 and single-input VOLO-D1 age & gender models accuracy. **Bold** indicates the best model for each benchmark. † marks the model that we release to the public domain.

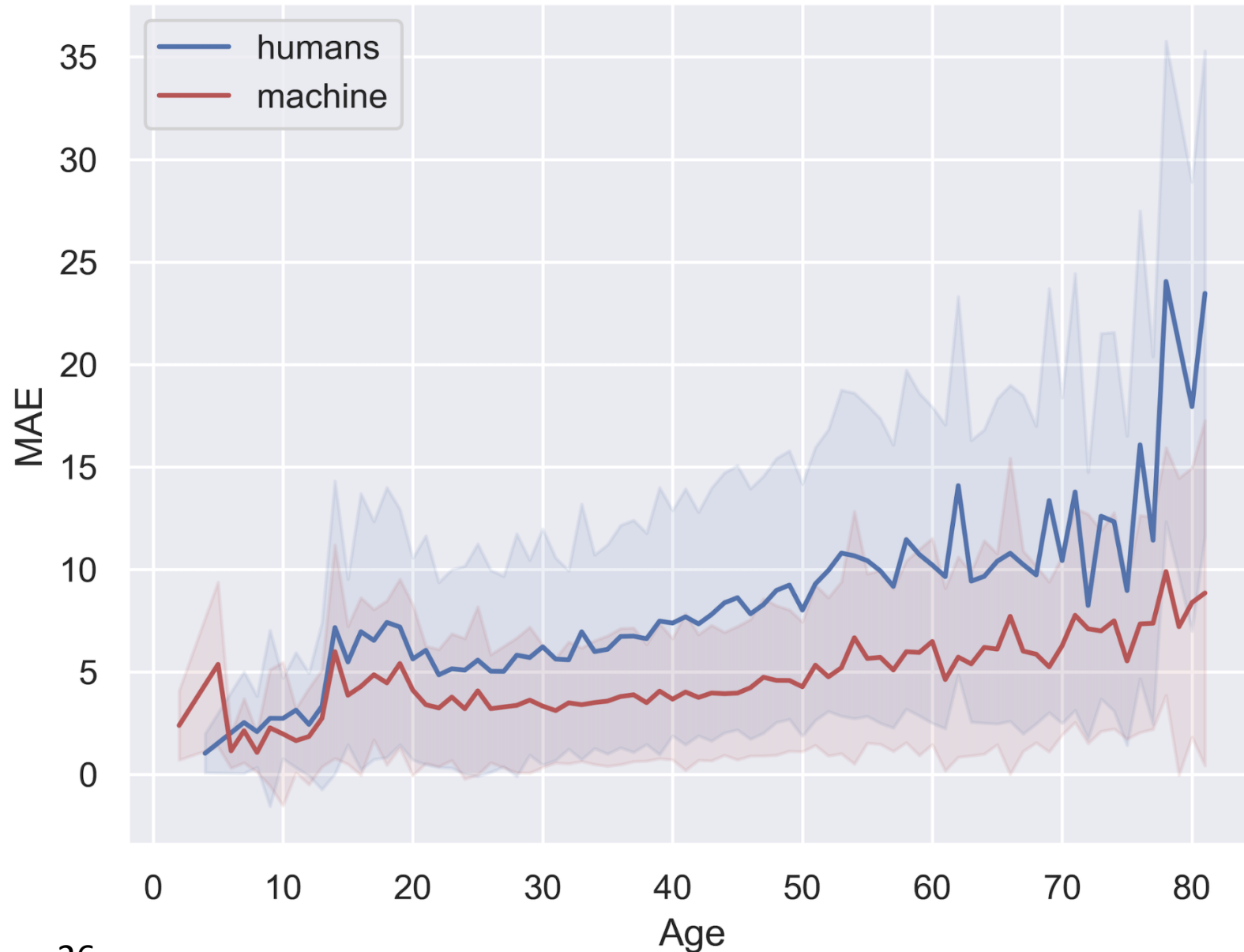


## Comparison with other models on LAGENDA

Method	Test Set	Age Acc	Age MAE	Gender Acc
FairFace[20] <b>MiVOLO-D1</b> <b>Face&amp;Body</b>	FairFace FairFace	59.70 <b>61.07</b>		94.20 <b>95.73</b>
DEX [33] [28] <b>MiVOLO-D1</b> <b>Face</b>	AgeDB AgeDB		13.1 <b>5.55</b>	- <b>98.3</b>
MWR [37] AL-ResNets-34 [43] Compacting [18] Gen MLP [22] <b>MiVOLO-D1</b> <b>Face</b>	Adience Adience  Adience Adience Adience	62.60 67.47  - - <b>68.69</b>		- -  89.66 90.66 <b>96.51</b>

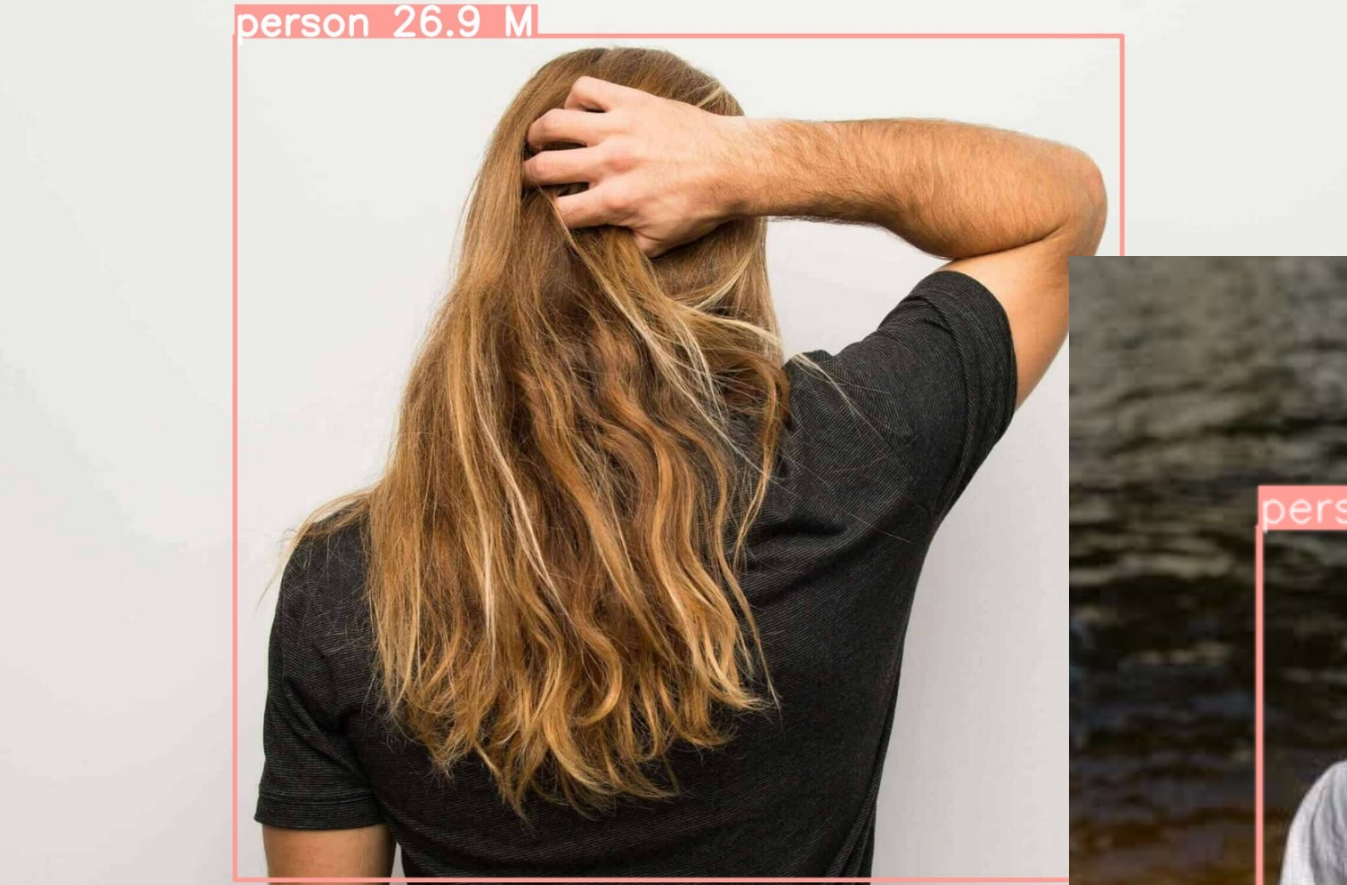
Table 4: FairFace, Adience, AgeDB validation results using MiVOLO-D1 trained on LAGENDA train set.

# Neural nets and human MAE comparison

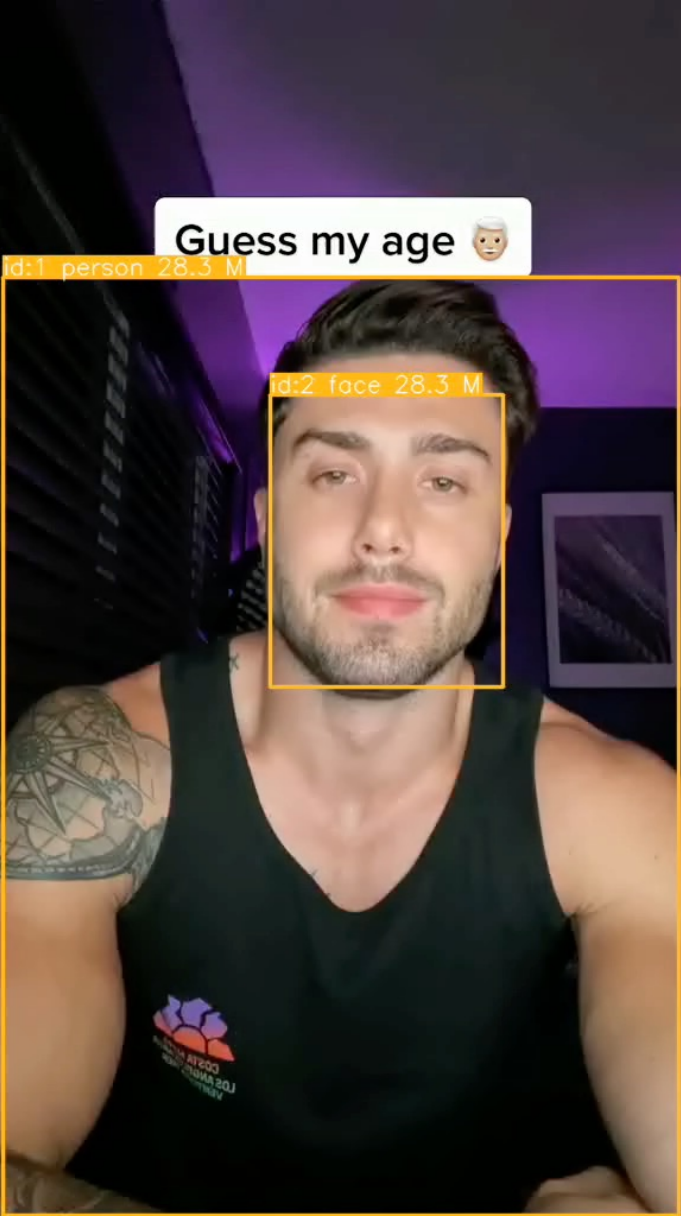


Scatter plot depicting the relationship between MAE and age on the IMDB-clean dataset, annotated by both human annotators and best MiVOLA.

# Examples



# Examples



# Examples



Looking for similar goods... ✓✓

Before



After



# Thank you for attention!

Code, demo, etc:

[https://github.com/](https://github.com/WildChlamydia/MiVOLO)

[WildChlamydia/MiVOLO](https://github.com/WildChlamydia/MiVOLO)

***Kuprashevich Maksim***

*mvkuprashevich@gmail.com*

*https://www.linkedin.com/in/wildchlamydia*

*https://t.me/wild\_chlamydia\_work*



***LayerCV TG: @layercv***