

# **Light Conditions Robustness Improvement in Skin Tone Estimation by Exploiting Picture Metainformation**

Dominik Agejev

Tara Baće

Mentor: Matija Šantek

PANTONE®  
P 45-11 C

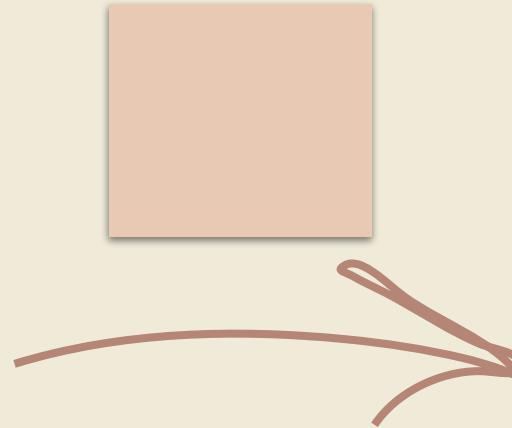
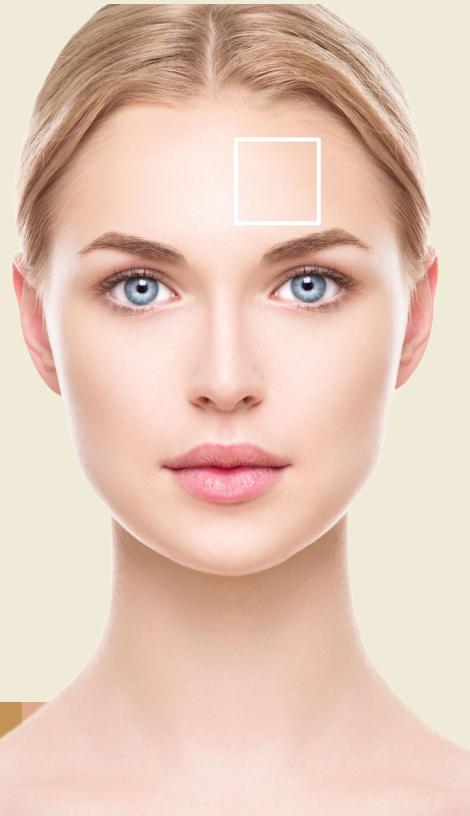
PANTONE®  
P 18-6 U

PANTONE®  
P 41-9 C

PANTONE®  
P 23-1 U

# arbelle

BY VISAGE TECHNOLOGIES



# Defining our Problem



- Lighting conditions
- Camera settings
- Phone preprocessing
- Image compression

**Captured color ≠ true color**

How to measure color correctly?

# $\Delta E$ Loss



- quantifies the **perceptual difference** between two colors
- Calculated in **CIELAB**

0-1	Almost imperceptible
1-2	Barely noticeable difference
2-3	Noticeable in close observation
3-6	Easily noticeable difference
6 <	Obvious difference

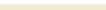
# Objective



- Robust color prediction under different lighting conditions using image **metadata**
- Estimate benefits for production use case

→ *Data preprocessing*  
→ *Feature engineering*

→ *Deep models*  
→ *Fusion techniques*



# Exif Metadata

## *Exchangeable Image File Format*

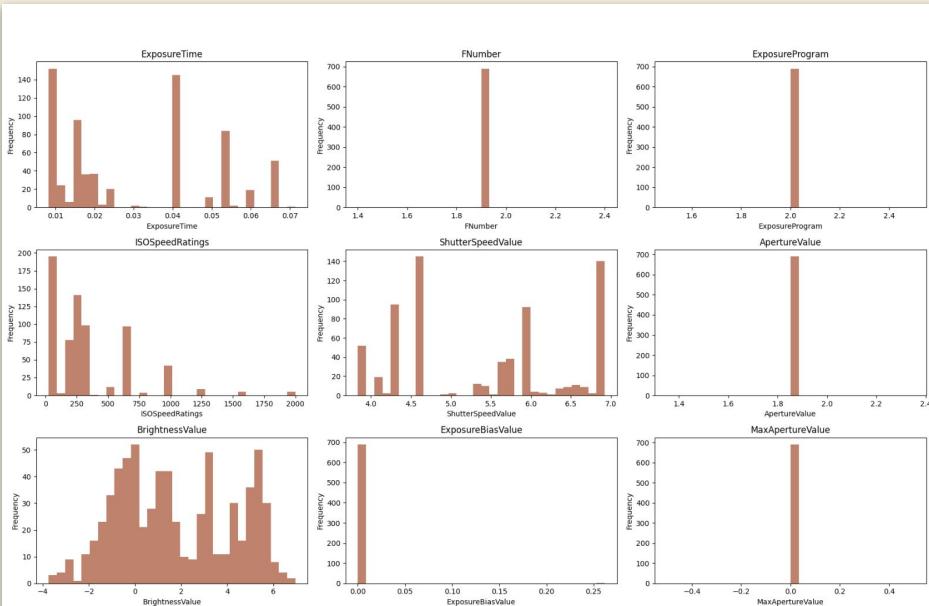
- Data automatically captured by digital camera
- Technical details, **exposure information**, location, ...

**ISO Speed Rating** measures the camera sensor's sensitivity to light.

**Shutter Speed Value** – logarithmic representation of Exposure Time.

Shows how much motion is captured in a photo and how much light was available.

**Brightness Value** – estimated brightness of the subject in focus.





138 cards



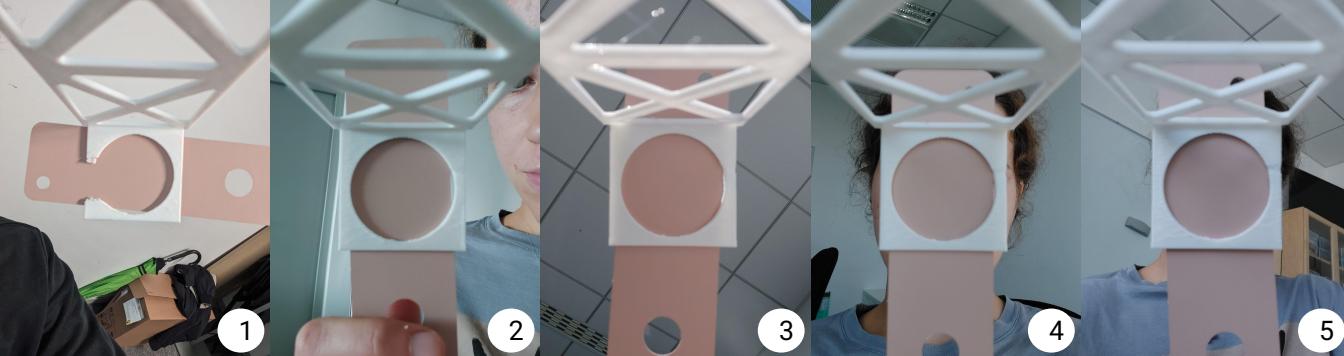
# Methodology

yellowbotM

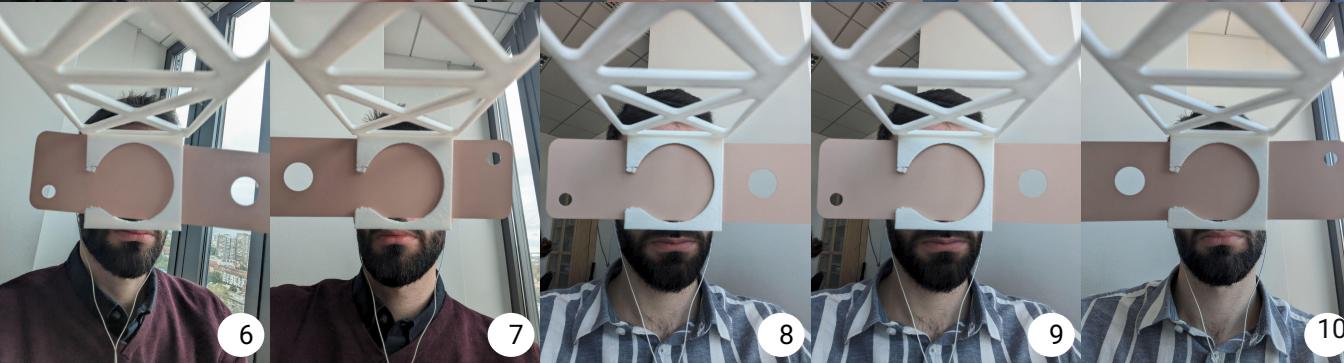


3 phones



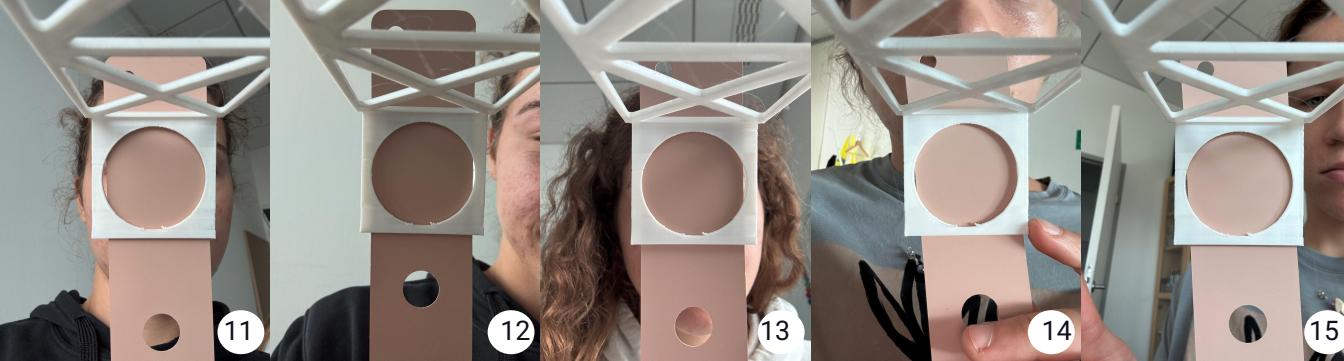


- 2070 images
- Validation: 15%
- Test: 20%
- (3 samples per color)



## 15 lighting conditions:

- 1) Ceiling light
- 2) Low light (hallway)
- 3) Bright light at an angle
- 4) Bright light
- 5) Natural light
- 6) Back light
- 7) Low backlight
- 8) Bright natural light
- 9) Natural front light
- 10) Natural light
- 11) Front light
- 12) Low back light
- 13) Cloudy
- 14) Low light (hallway)
- 15) Natural light



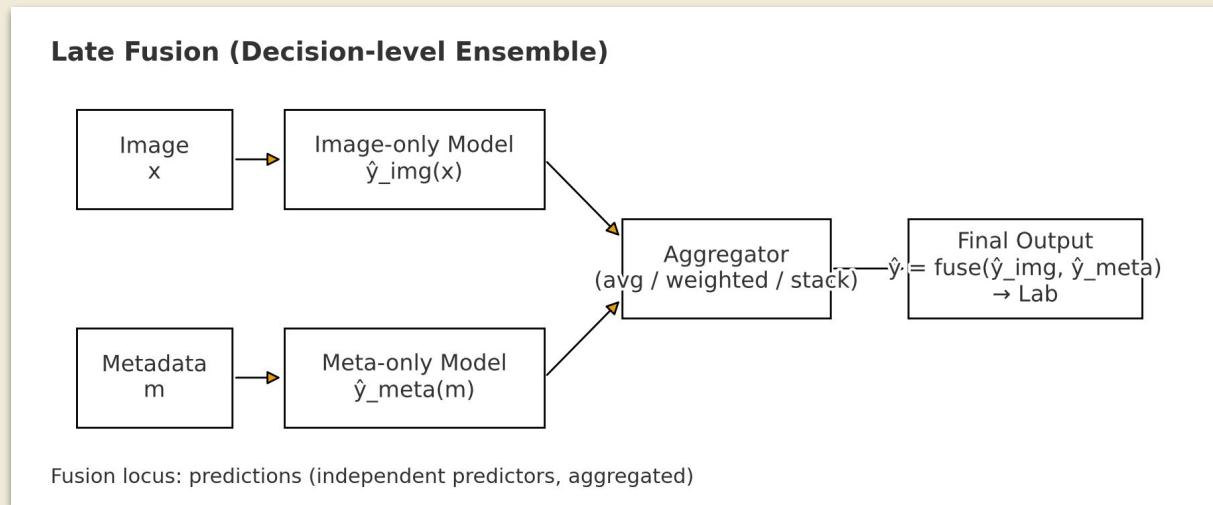
# Data Overview



	image	image + mean	image + meta	image + meta + mean
cropped				
background				

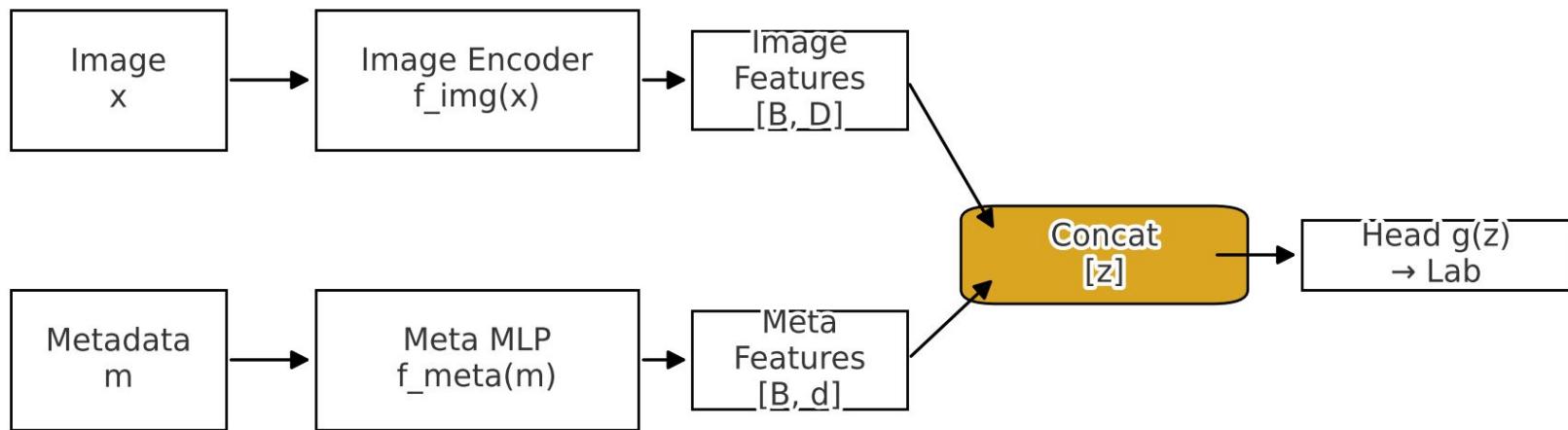
# Fusion

- How to combine image and text data?
- Early fusion – data level
- **Middle fusion – feature level**
- Late fusion – decision level





## Feature-level Fusion (Concatenation)

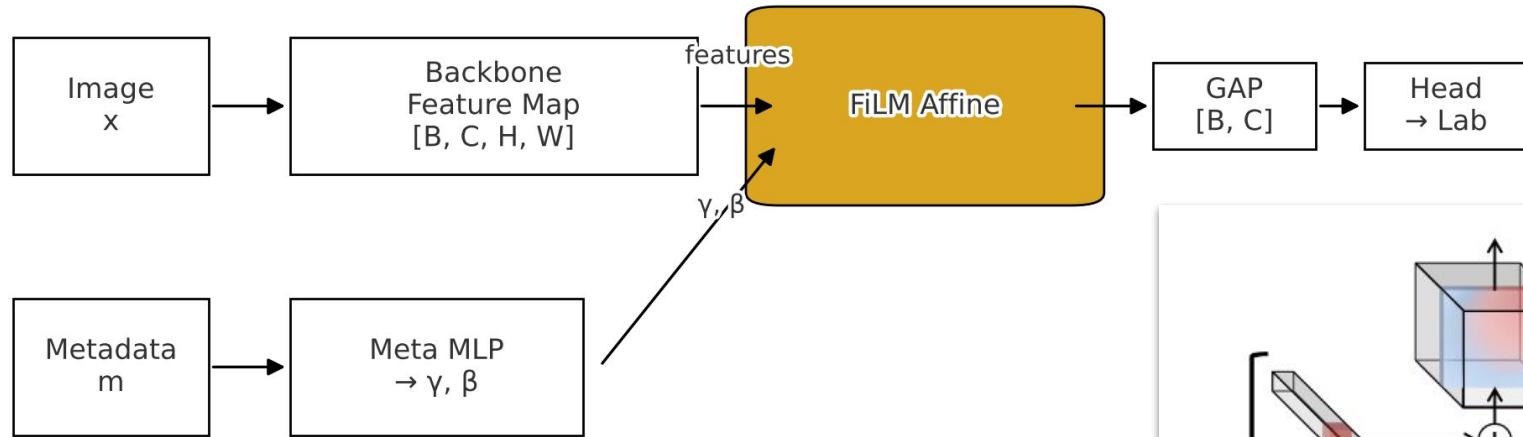


Fusion locus: feature vectors (intermediate)

- encode image and metadata separately → concatenate features → head → Lab

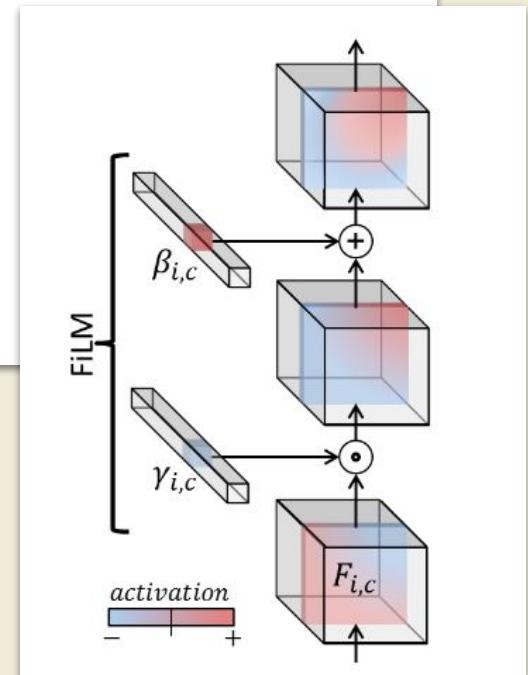


## FiLM: Metadata-Conditioned Features



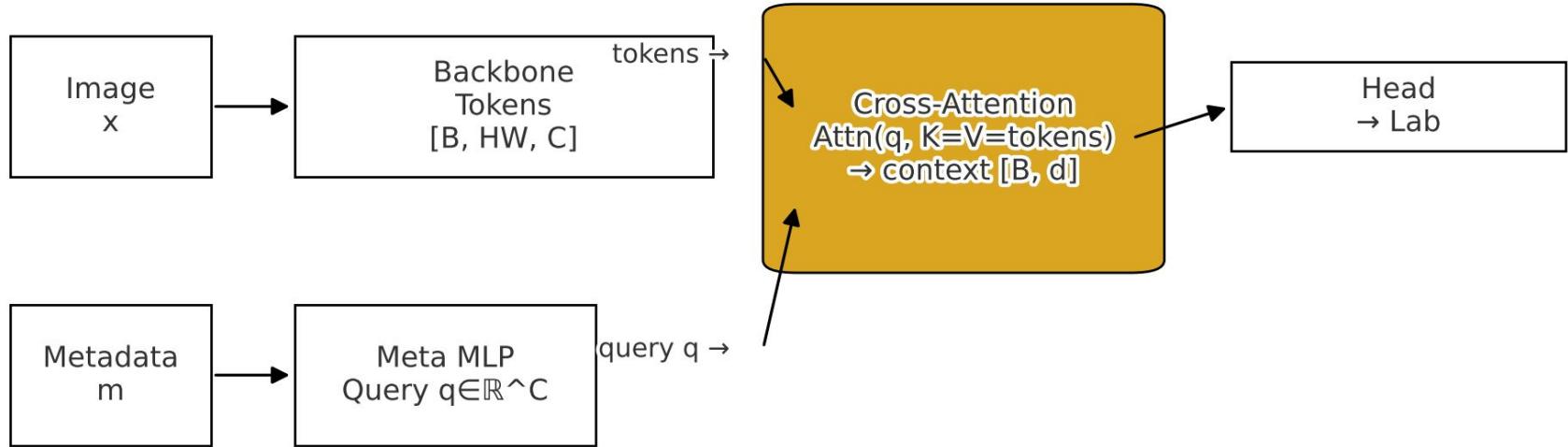
- Feature-wise Linear Modulation:

$$X^{\square'} = (1 + \alpha \cdot \tanh(\gamma^{\square})) \odot X^{\square} + \alpha \cdot \tanh(\beta^{\square})$$





## Cross-Attention: Metadata Queries Image Tokens

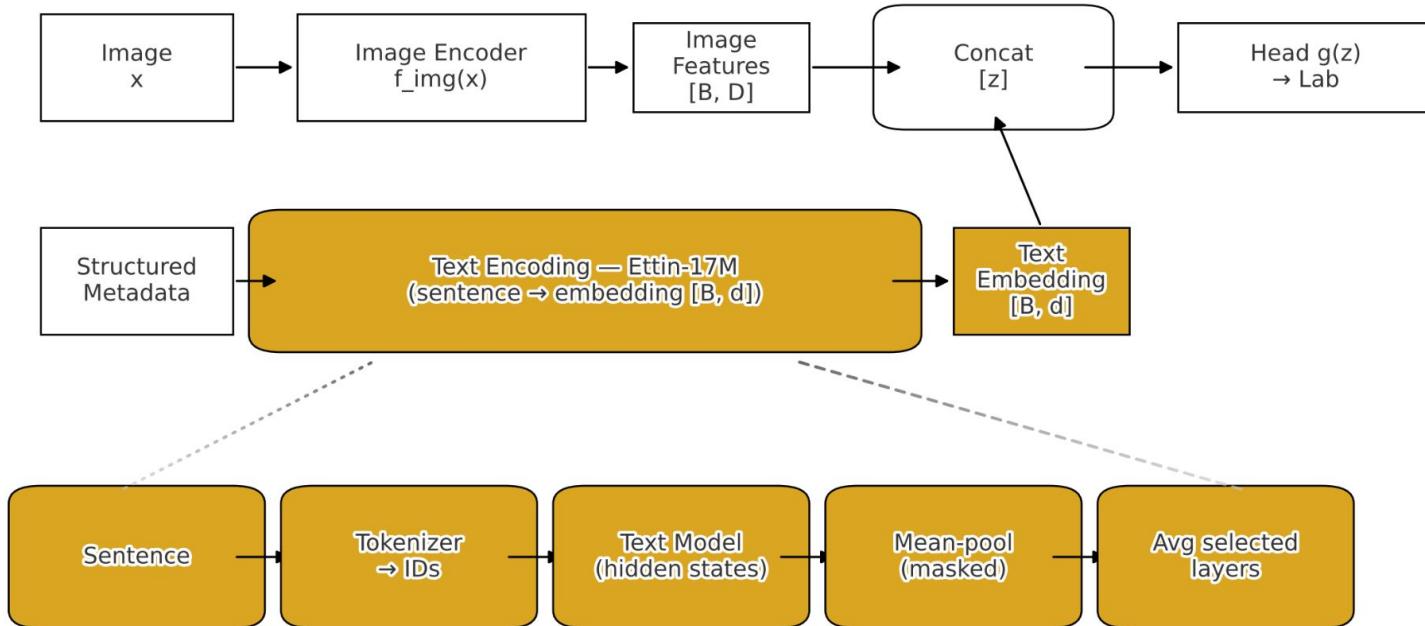


Fusion locus: token/feature level via attention

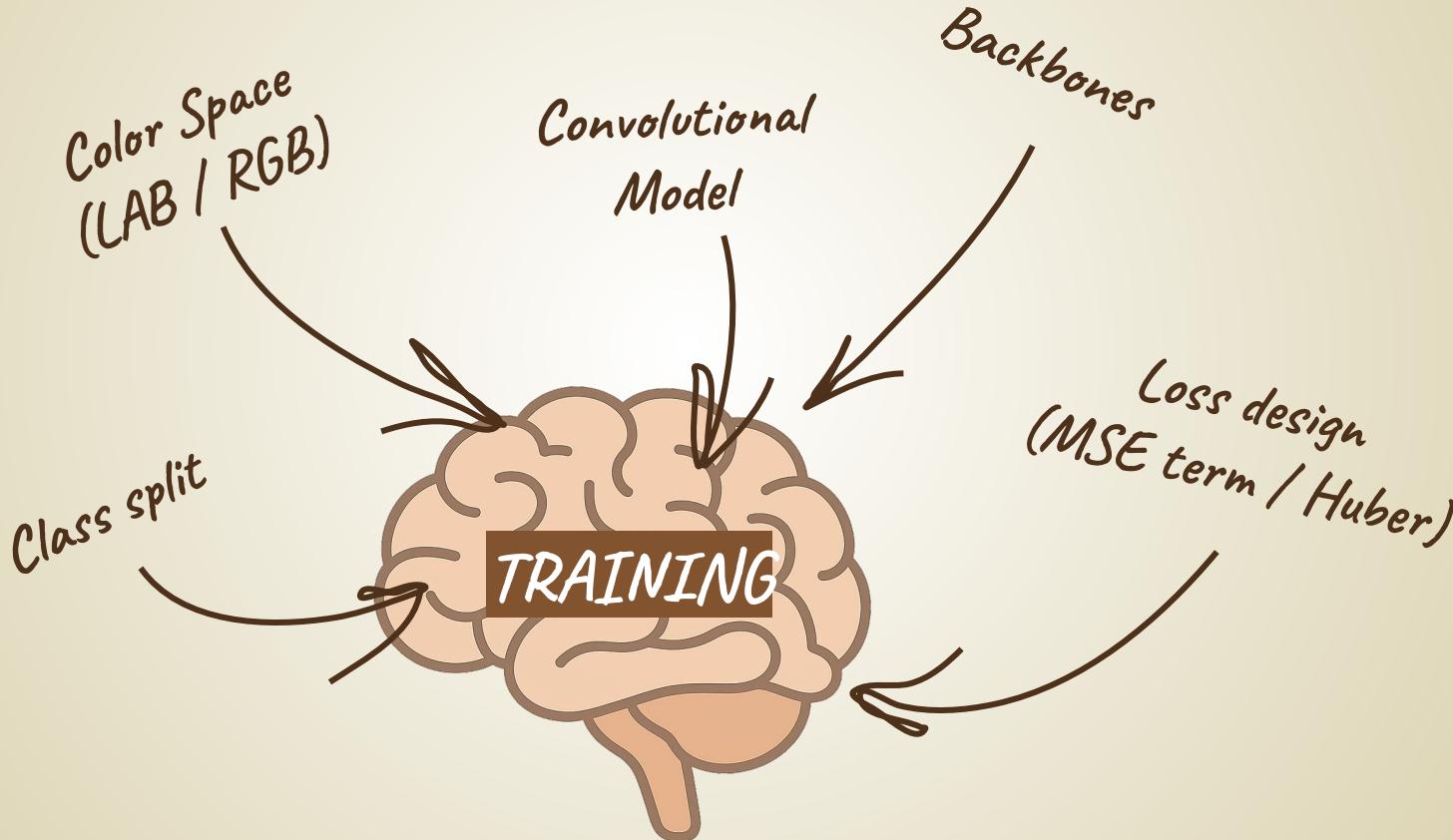
- backbone → image tokens; metadata → query;  $\text{Attn}(q, K=V=\text{tokens}) \rightarrow \text{context} \rightarrow \text{head.}$



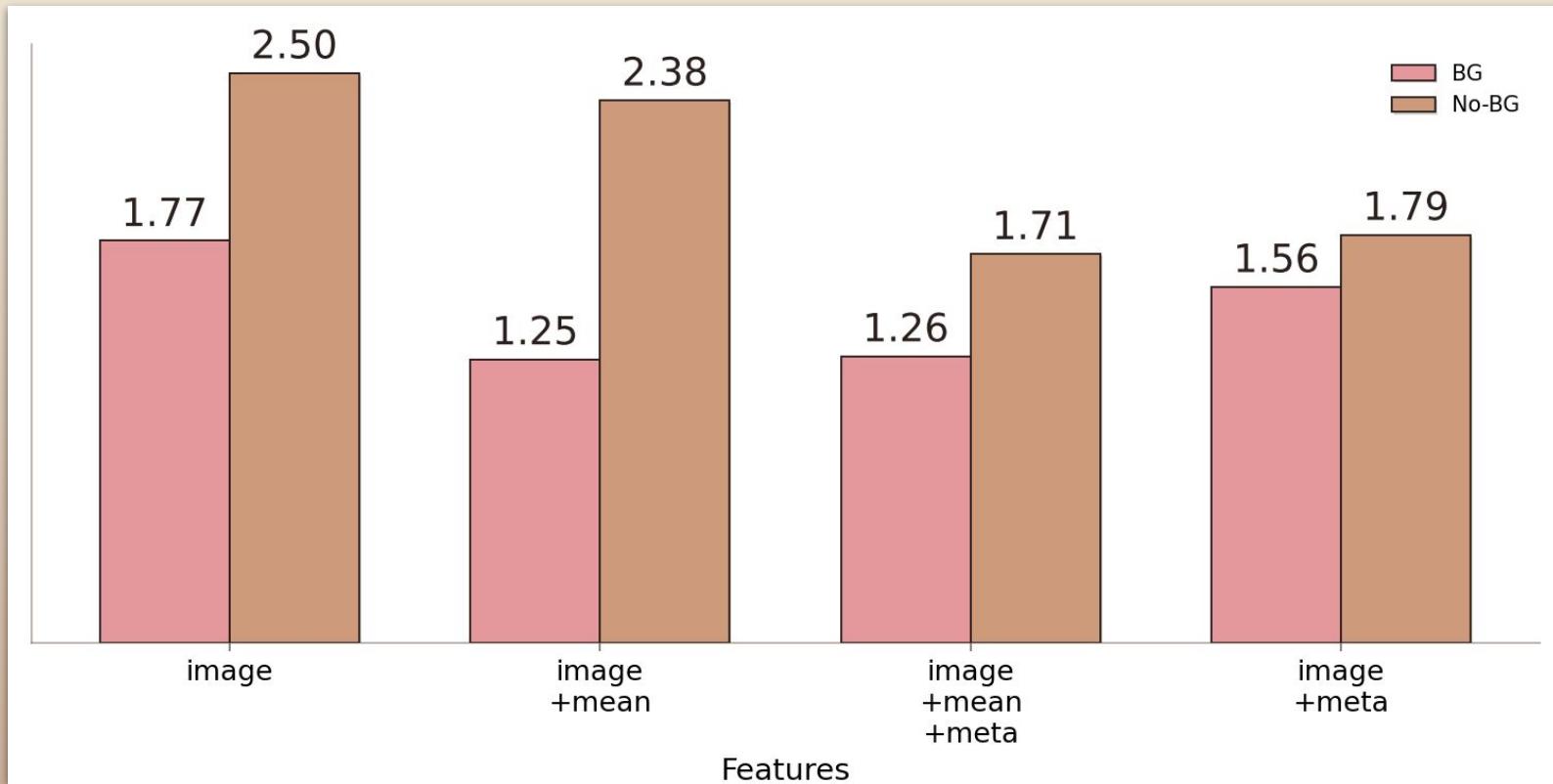
## Feature-level Fusion (Concat) + Text Encoder



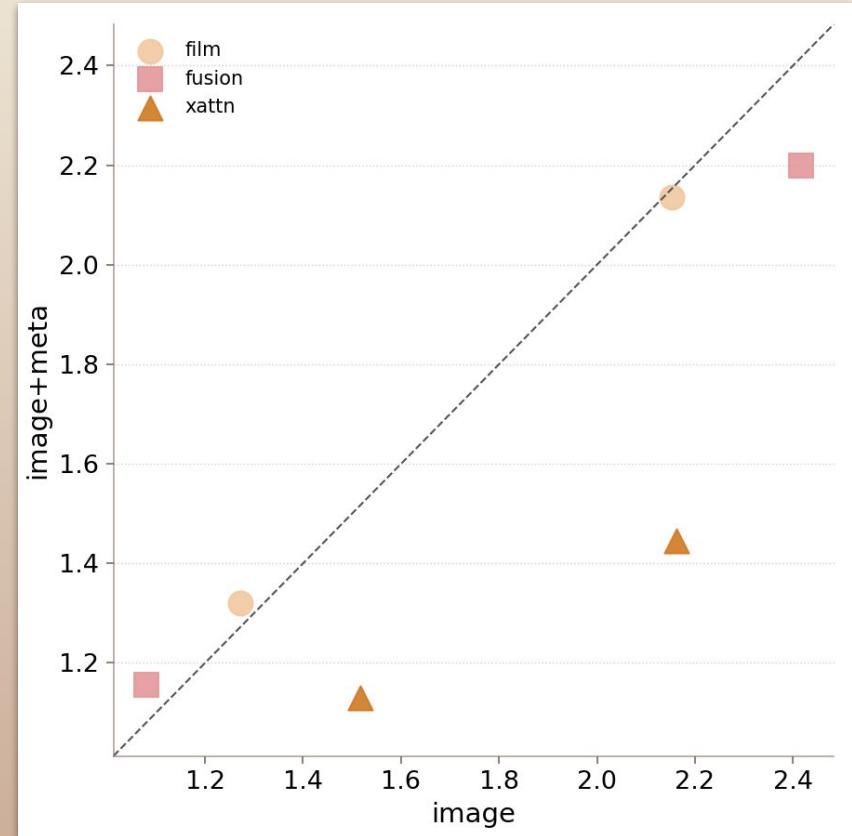
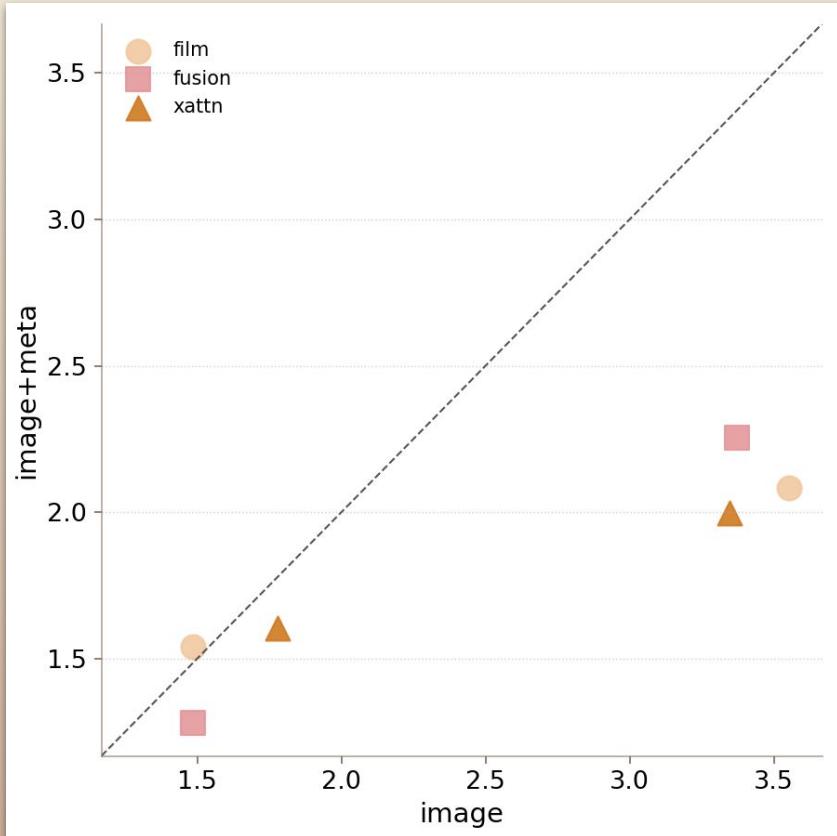
Example sentence: iso log 4.20; shutter log -6.11; ev +0.42; wb auto; make Apple



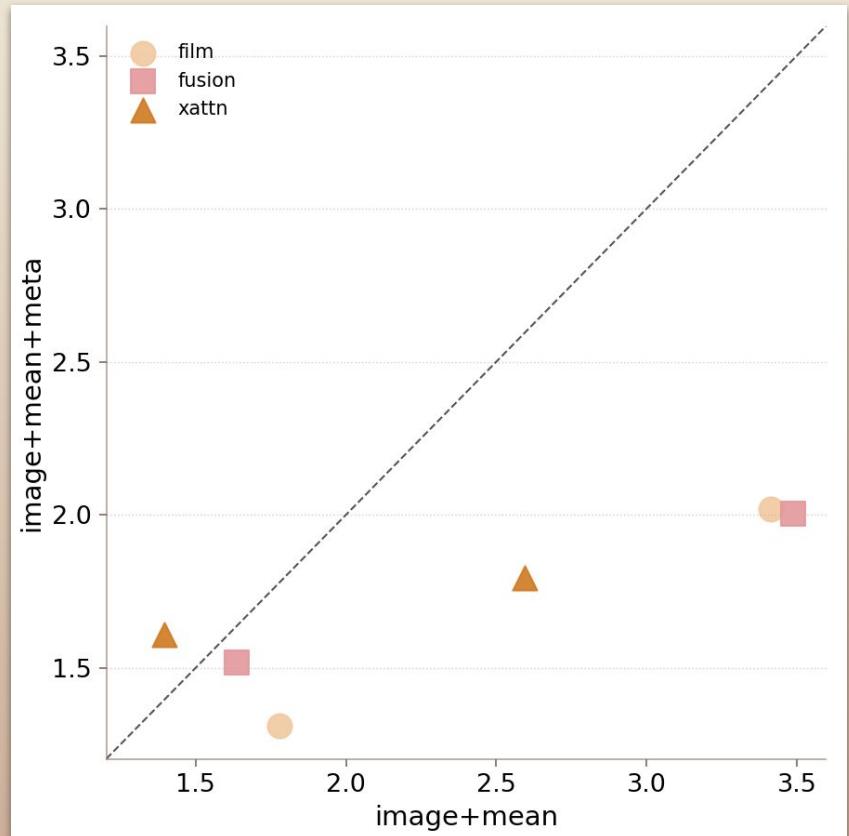
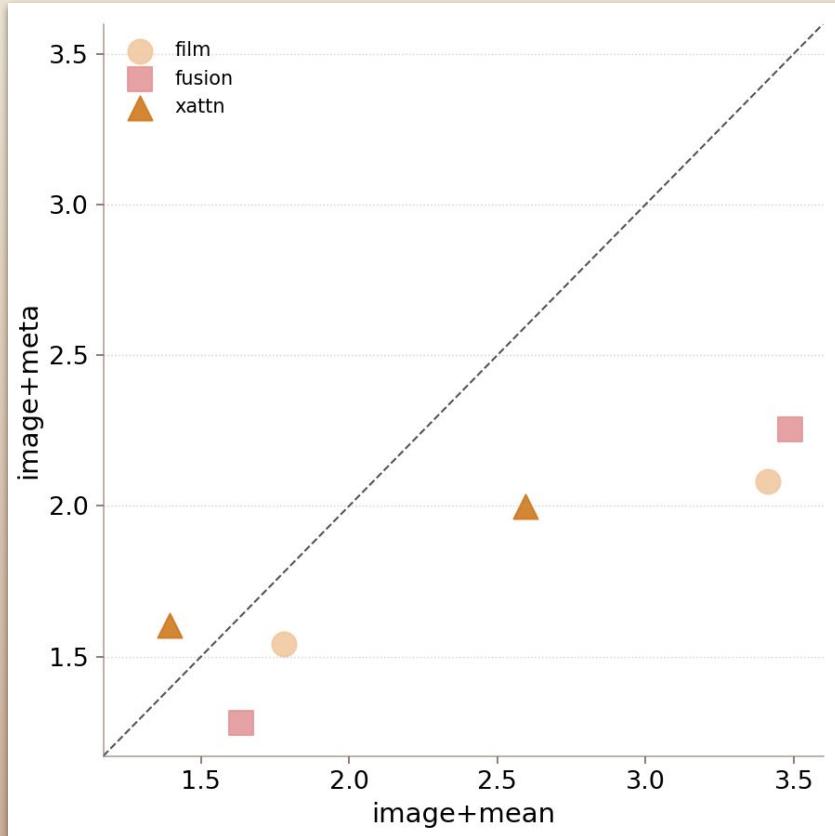
# Results by features • background



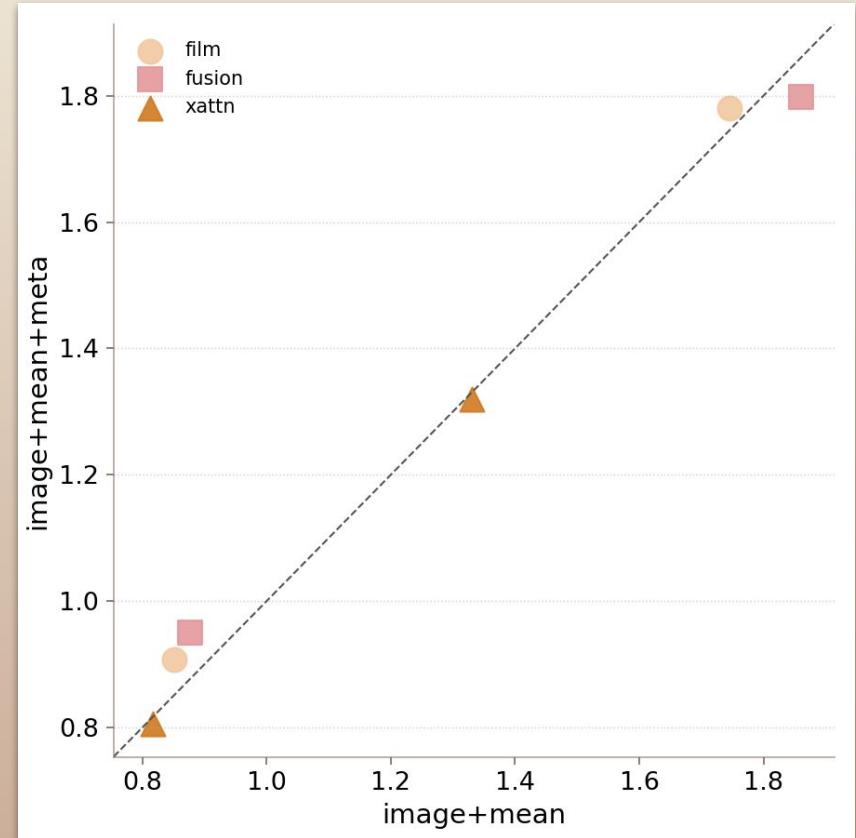
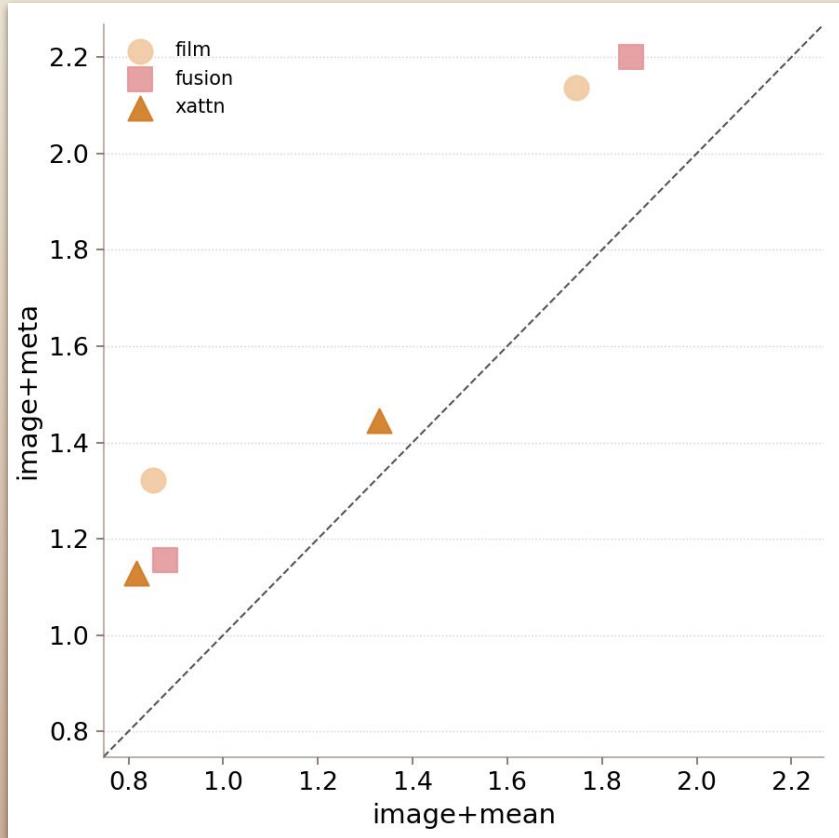
# Metadata *without* vs. *with* background



# Metadata vs. mean *without background*



# Metadata vs. mean *with background*



# Conclusion



- **Data is king**
  - Focused images
  - Adaptive cropping
  - Merged with background
- **Metadata – good not great**
  - Cropped images + metadata → significantly better performance
  - Background context reduces the added value of metadata
  - Mean feature outperforms metadata
  - Has potential for production use case



# Next steps



PANTONE® P 25-1 U PANTONE® P 41-9 C PANTONE® P 18-6 U PANTONE® P 45-11 C PANTONE® P 50-5 U PANTONE® P 19-1410 TSK Chocolate Drizzle PANTONE® P 25-1 U PANTONE® P 41-9 C PANTONE® P 18-6 U PANTONE® P 45-11 C PANTONE® P 50-5 U PANTONE® P 19-1410 TSK Chocolate Drizzle

- End-to-end trainable encoder
- Geometric and photometric data augmentation
- Original architecture / multi-stage cross-attention
- Late fusion



PANTONE® P 25-1 U PANTONE® P 41-9 C PANTONE® P 18-6 U PANTONE® P 45-11 C PANTONE® P 50-5 U PANTONE® P 19-1410 TSK Chocolate Drizzle

