MVIP2024

13th Iranian and 3rd International Conference on Machine Vision and Image Processing

6 & 7 March , 2024

MVIP2024.KHU.AC.IR

Cultural-aware AI model for emotion recognition

Mehrdad Baradaran
Shahid Beheshti University

Payam Zohari
Khajeh Nasir Nniversity of Technology

Abtin Mahyar
Institute of Research in Fundamentals

Hossein Motamednia
Institute of Research in Fundamentals

Dara Rahmati
Shahid Beheshti University

Saeid Gorgin
Chosun University

Presentation Outline:

- Problem definition
- Dataset introduction
- Proposed methods
  - Image Model
  - Text Model
- Linear Combination Model
- Results
- Conclusion

Do cultures affect our interpretation?

How to utilize captions to Convey emotions?

WIKIART

Construction of an emotional image captioning model, requires the task of image processing as well as emotion recognition. The “WIKIART” dataset, supplying thousands of classified paintings helped us through this journey.

ArtELingo

To approach cultural-awareness model, we needed to tra
our model on a dataset rich in multilingual captions an
“ArtELingo” provided us with a collection of 80k
annotated artworks in 3 languages, namely English,
Arabic, and Chinese.

But why “ArtELingo”?

	COCO	ArtEmis	ArtELingo
Image Source	Photos	WikiArt	WikiArt
#Images	328k	80k	80k
#Annotations	2.5M	0.45M	1.2M
#Annot/Image	7.6	5.68	15.3
Emotions	0	9	9
Languages	E	E	ACES

Proposed Method

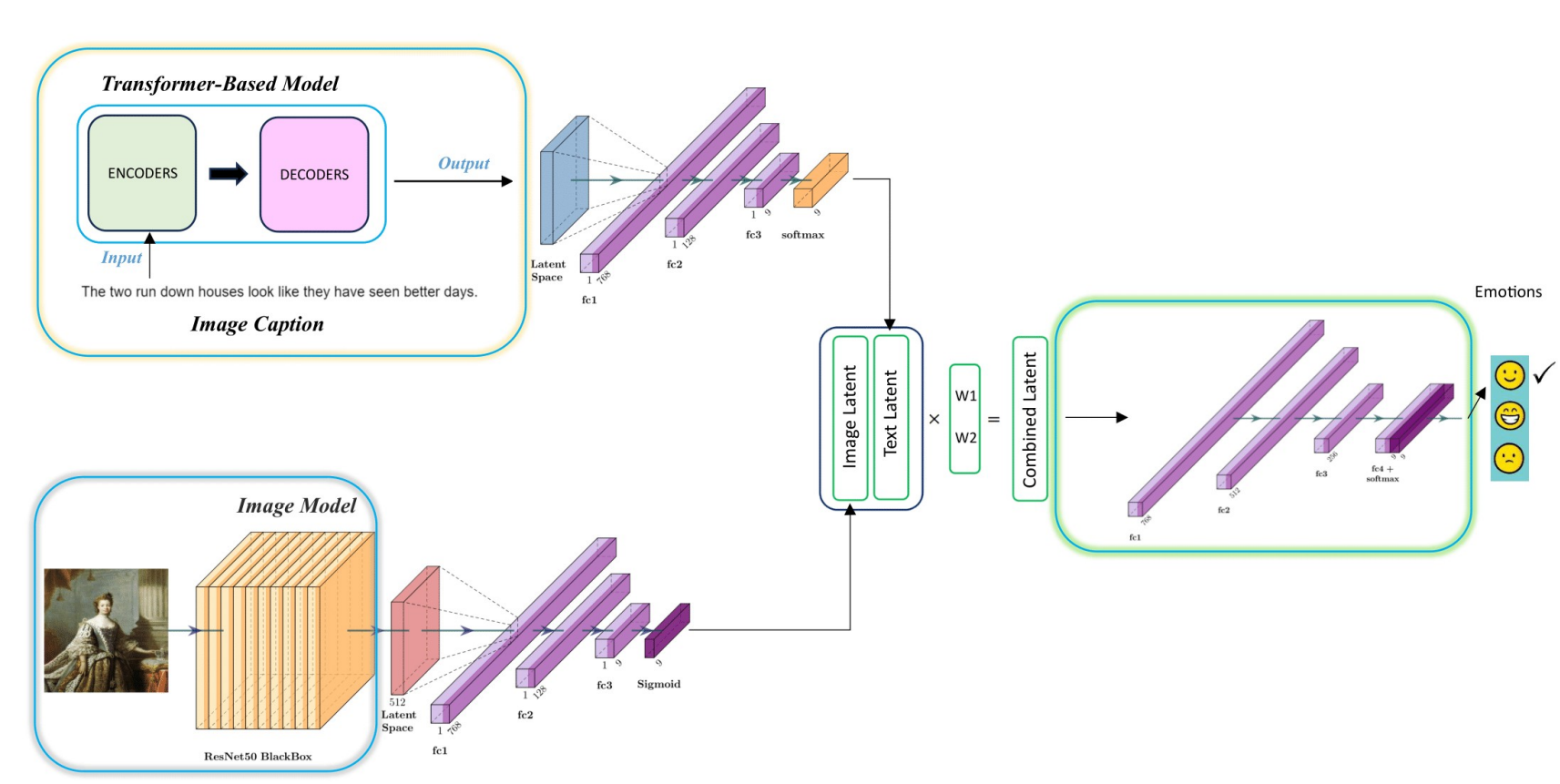
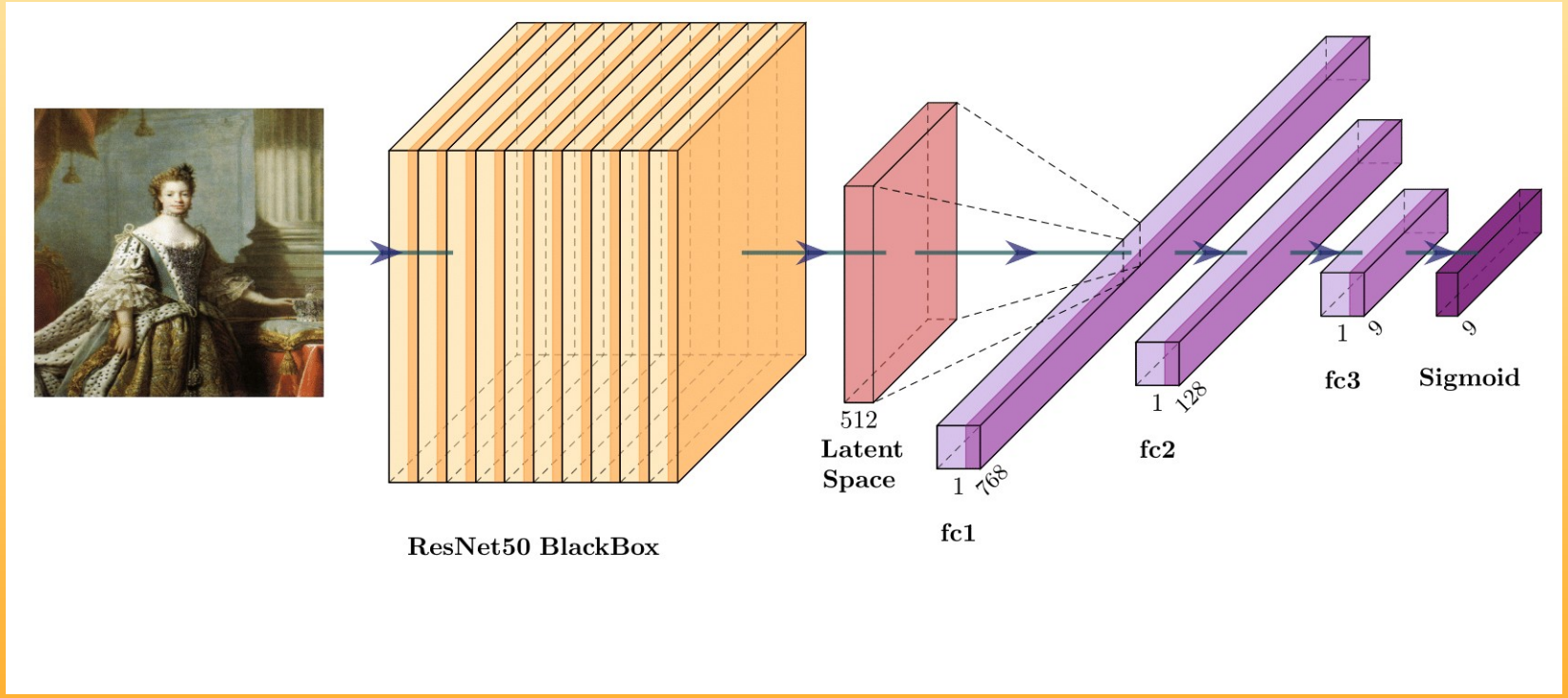
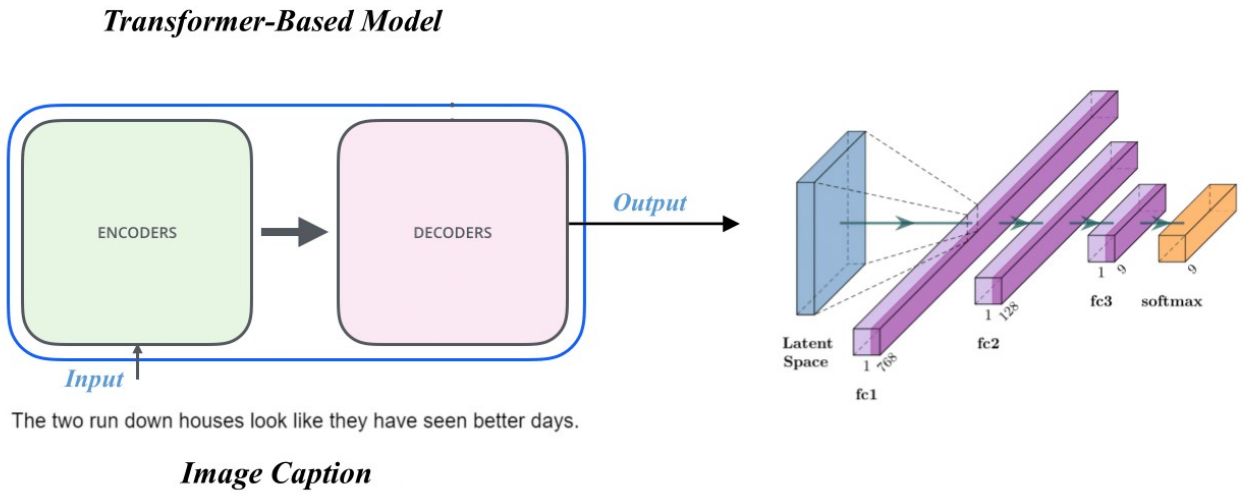


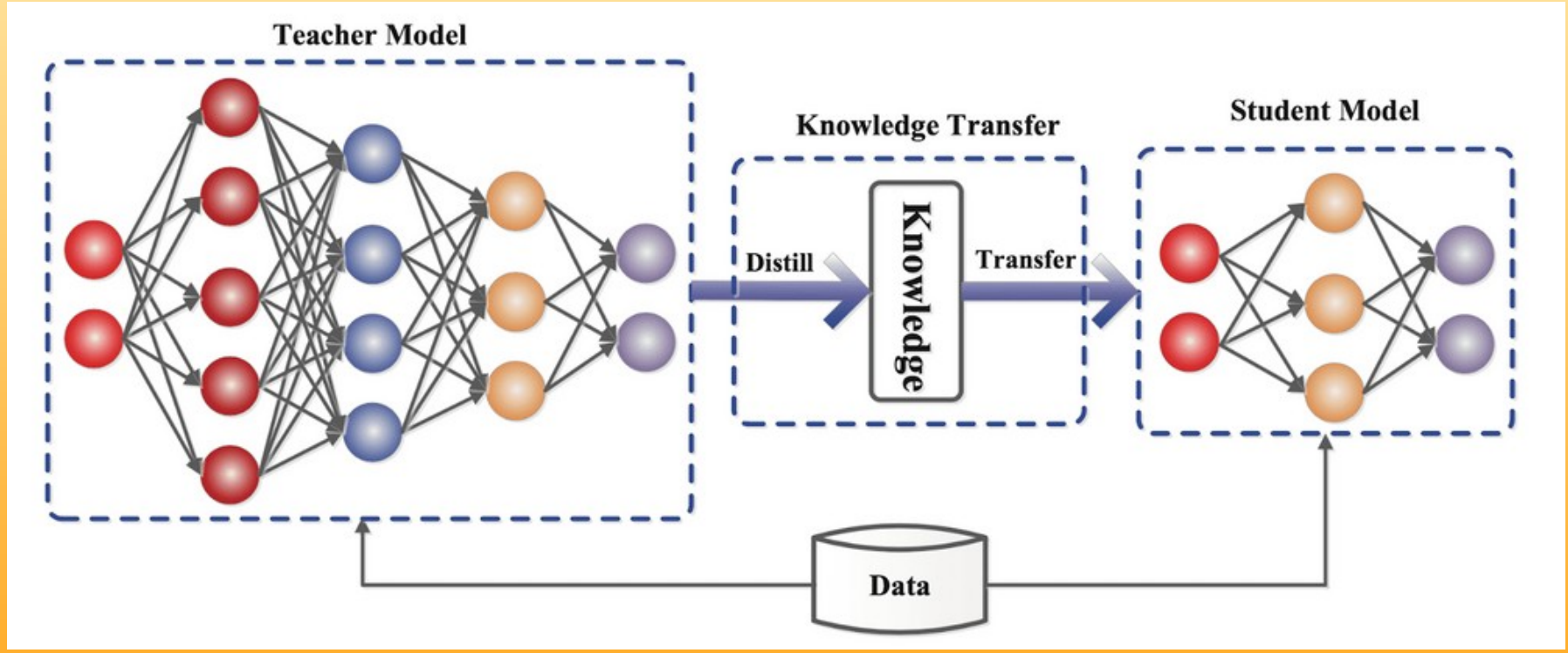
Image Model


ResNet50 BlackBox
512
Latent
Space
1
768
fc1
1
128
fc2
1
9
fc3
9
Sigmoid

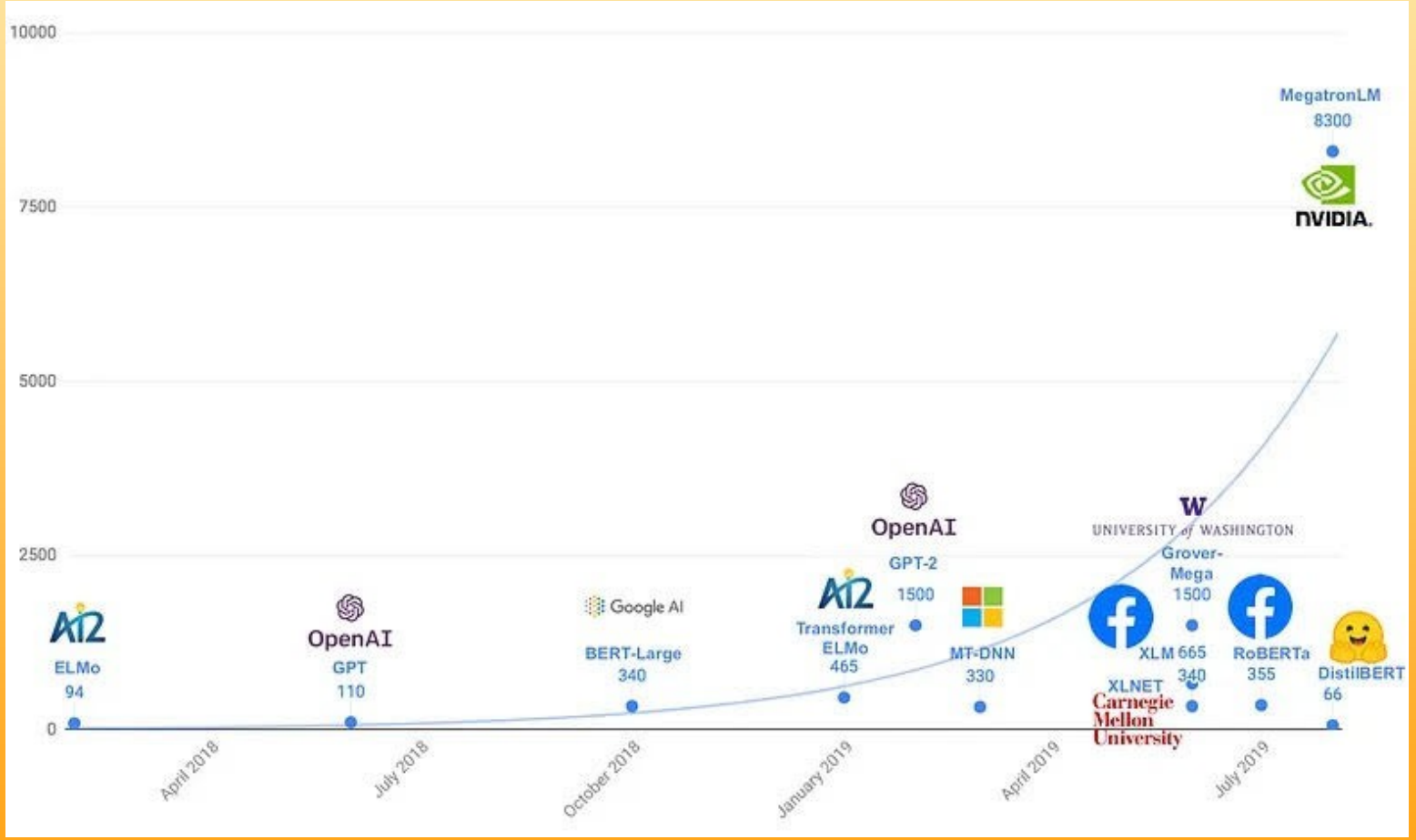
Text Model


Transformer-Based Model
ENCODERS
DECODERS
Output
Input
The two run down houses look like they have seen better days.
Image Caption
Latent Space
1 768
fc1
1 128
fc2
1 9
fc3
9
softmax

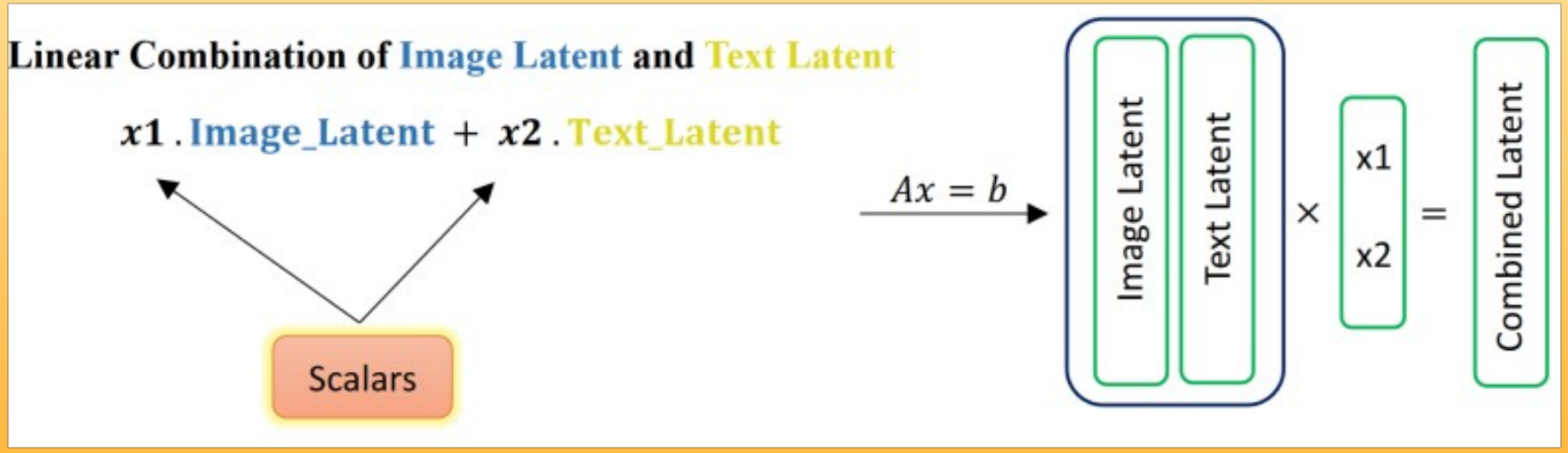
Knowledge Distillation



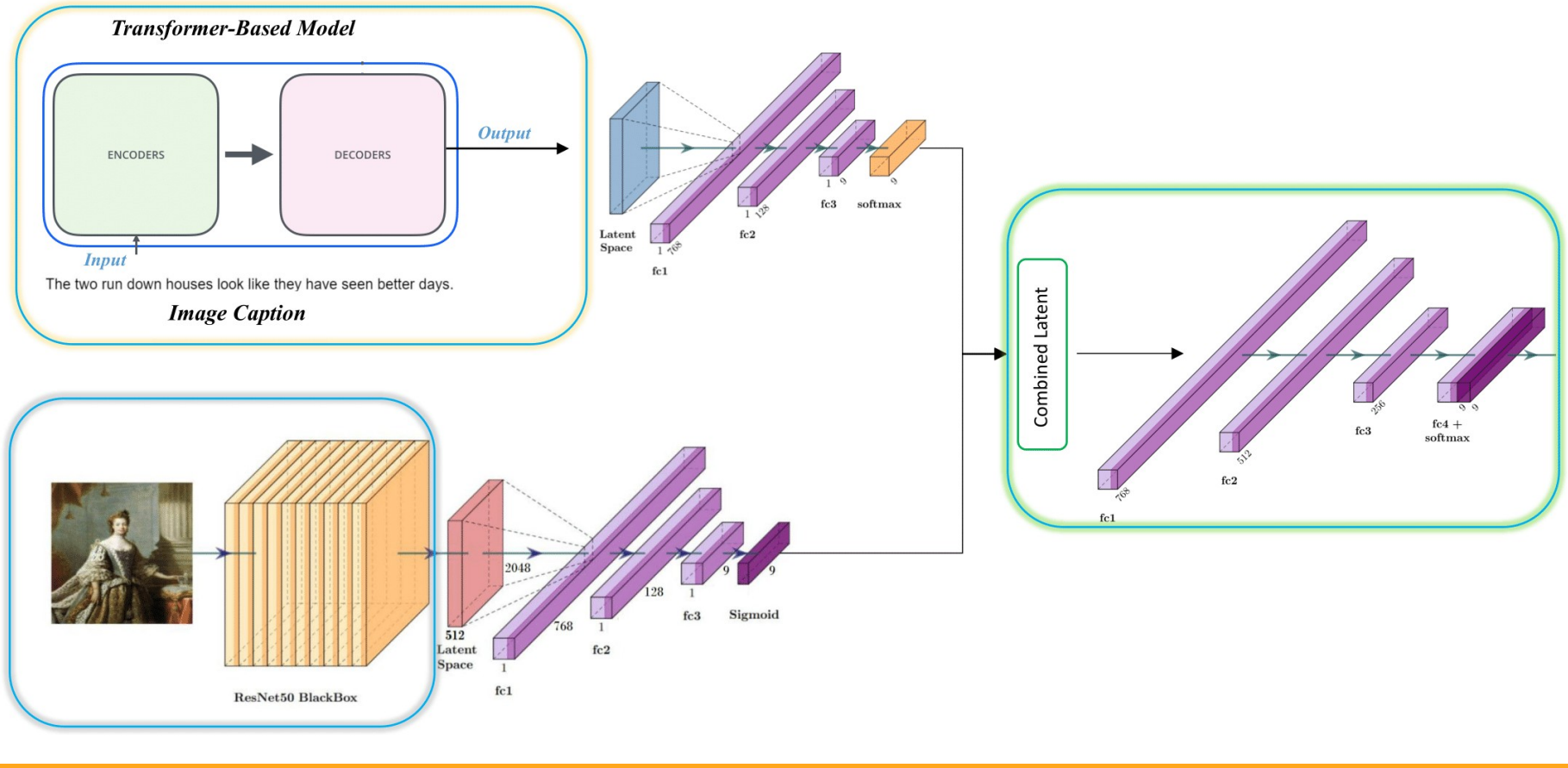
Why DistilBERT?



Feature
Linear Combination



Model Overview

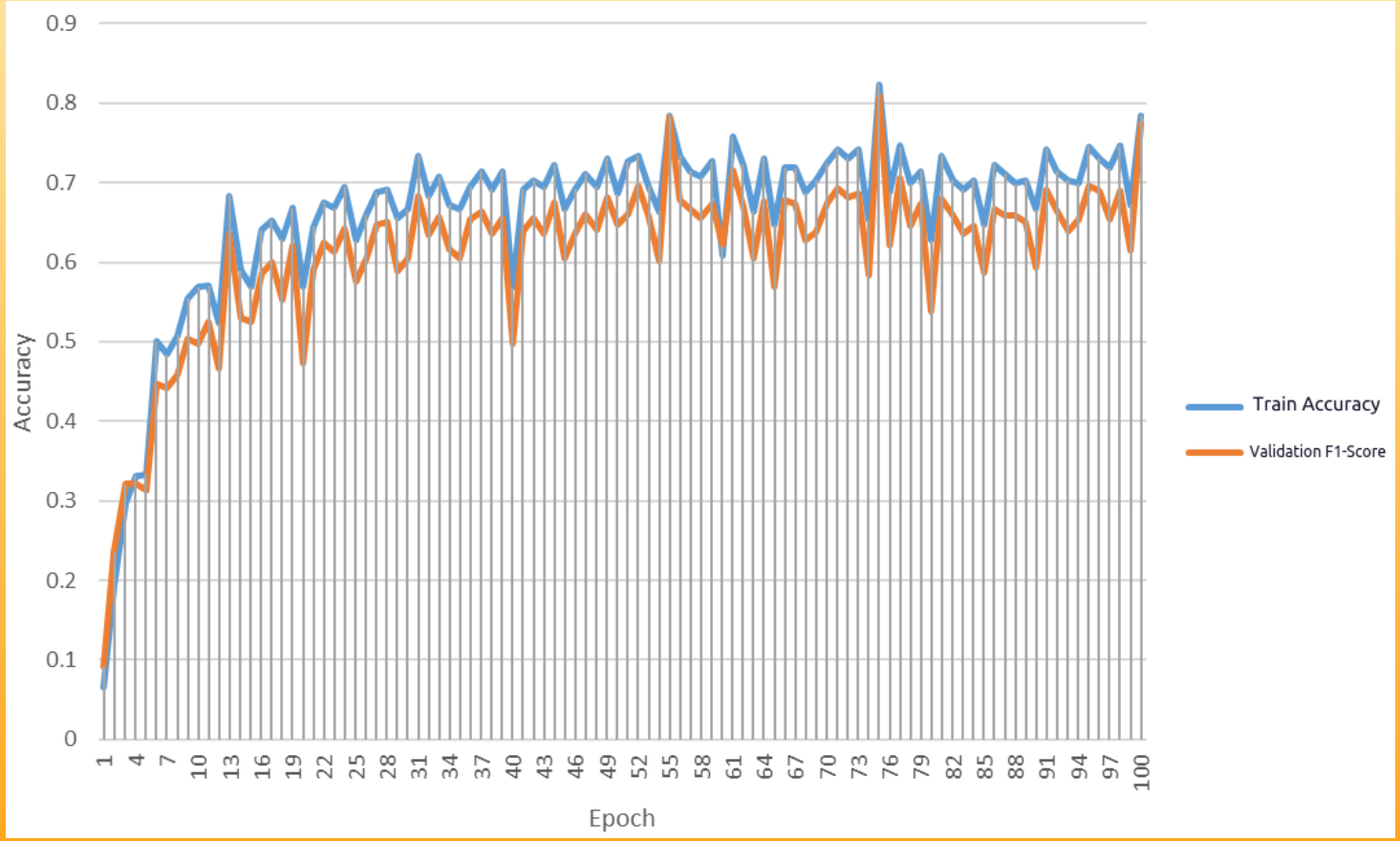


Our Results

THE EVALUATION OF THE IMAGE MODELS ON THE ARTELINGO DATASET

Model/Accuracy	Train ACC	Validation Acc
ResNet50	0.711386	0.698935688
VGG	0.717367	0.706483998
VIT	0.702534	0.687801932

Our Results


0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Accuracy
1
4
7
10
13
16
19
22
25
28
31
34
37
40
43
46
49
52
55
58
61
64
67
70
73
76
79
82
85
88
91
94
97
100
Epoch
Train Accuracy
Validation F1-Score

Thanks for your attention!