Machine Learning Course Project

Acne Counting and Grading

Group 17

2050735 Li Hengxin, 2151214 Mi Tiantian,
2152050 Rao Ji, 2152667 Li Ao

Part 1. Background and Motivation

- **Acne**:

  A common chronic inflammatory skin disease, often found in teenagers.

- **Effects of acne on daliy life**:

  Acne is often associated with stress, staying up late, irregular diet and other bad habits, seriously affecting the appearance and quality of life of patients.

- **Treatment method**:

  Doctors need to precisely target acne for topical treatments or laser therapy, both of which require accurate localization of the acne.

Part 1. Background and Motivation

- **Why we need acne localization**?

  In clinical practice, it is <mark>laborious</mark> for dermatologists to diagnose acne grade manually. The purpose of this project is to use the method of object detection in deep neural network to solve it.
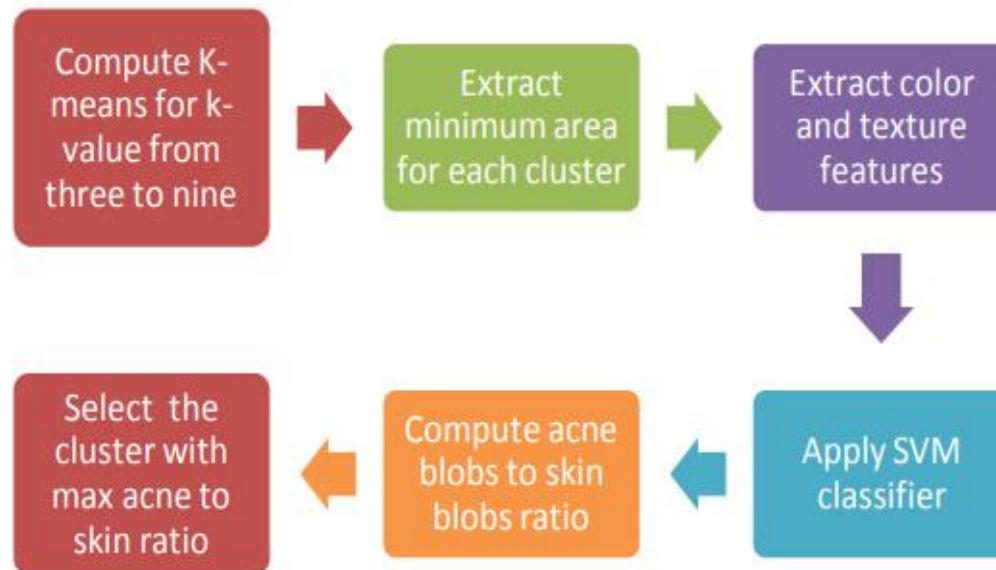
- **Practical significance**:

  - Enhancing doctors' <mark>efficiency</mark>

  - Provide services to patients doctors

  - Offer more effective <mark>treatment guidance</mark>

Fig. 1. The example of four lesion (acne) images. From left to right: Comedone, Papule, Pustule, Nodule.
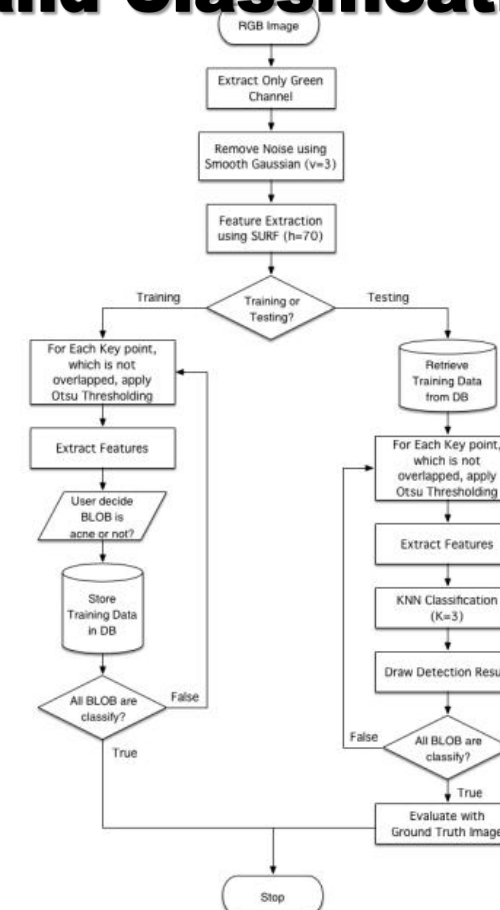
- **Conventional approaches**:
  suffer from noise and low accuracy

Digital Assessment of Facial Acne Vulgaris
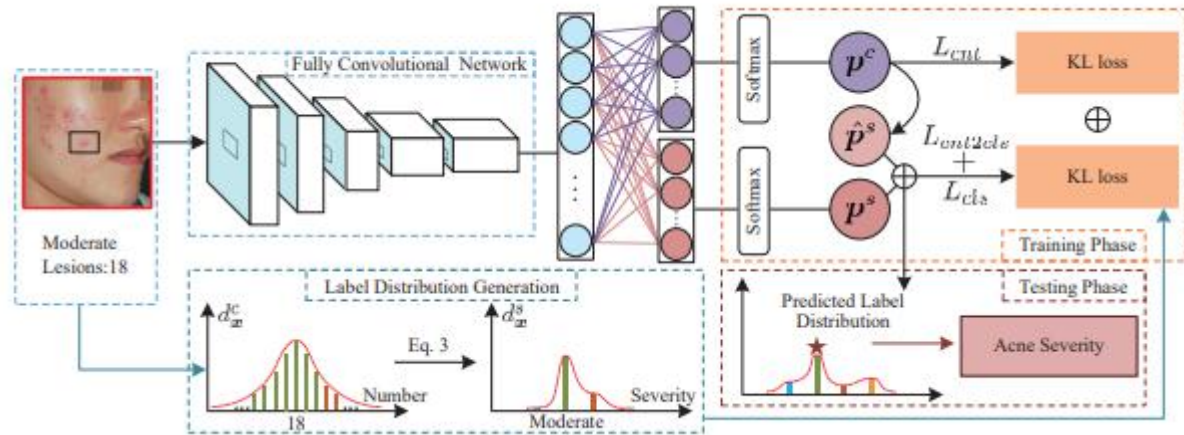(2014 I2MTC Proceedings)

Pipeline: Manual feature + SVM

Acne Detection Using Speeded up Robust Features and
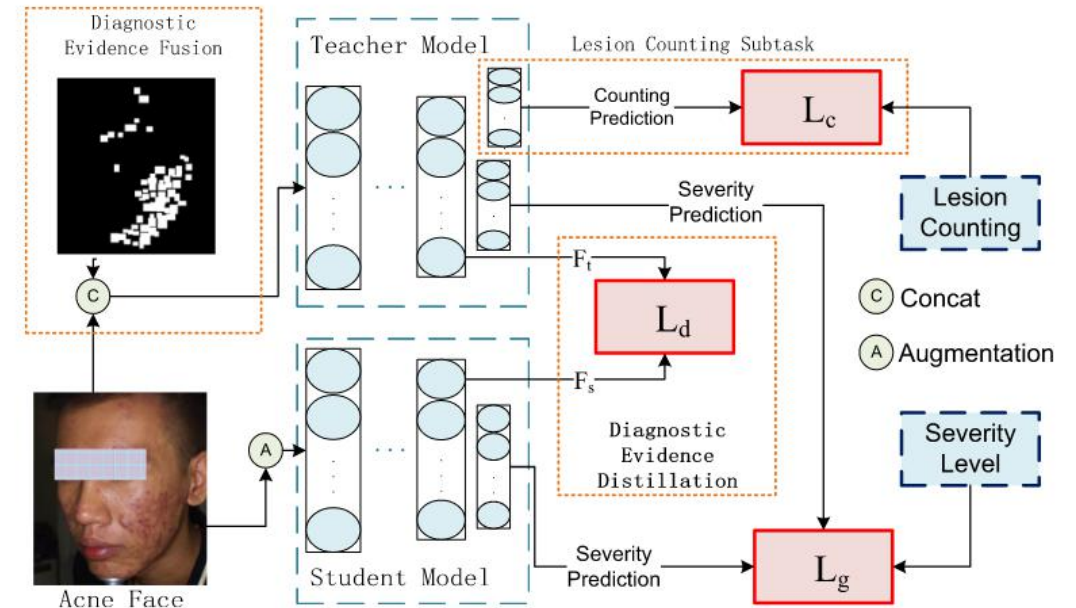Quantification Using K-Nearest Neighbors Algorithm (2014 ICCBS)

Pipeline: SURF + KNN

- **Deep learning-based methods**:
  Regression, classification and segmentation tasks have many applications in medical image analysis.

Joint Acne Image Grading and Counting via Label Distribution Learning (2019 ICCV)

DED: Diagnostic Evidence Distillation for acne severity grading on face images (2023 Expert Systems With Applications EI)

Pipeline: Counting Distribution Severity Distribution in ResNet Backbone Generating Label Distribution

Pipeline: knowledge distillation framework incorporates with joint learning for the teachernetwork

Part 2 Related Works : Object Detection

YOLOV5s Struture
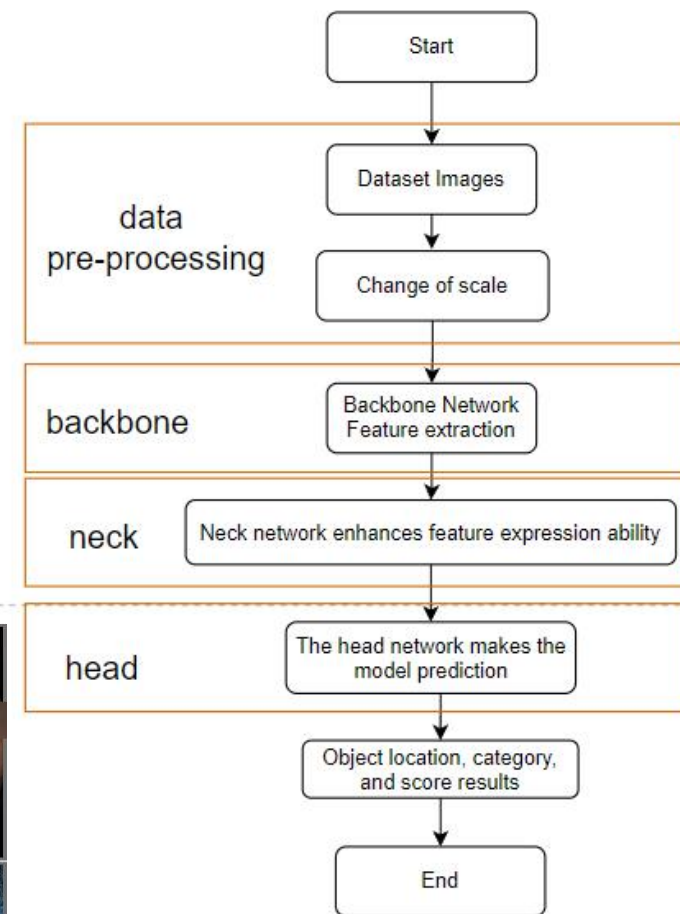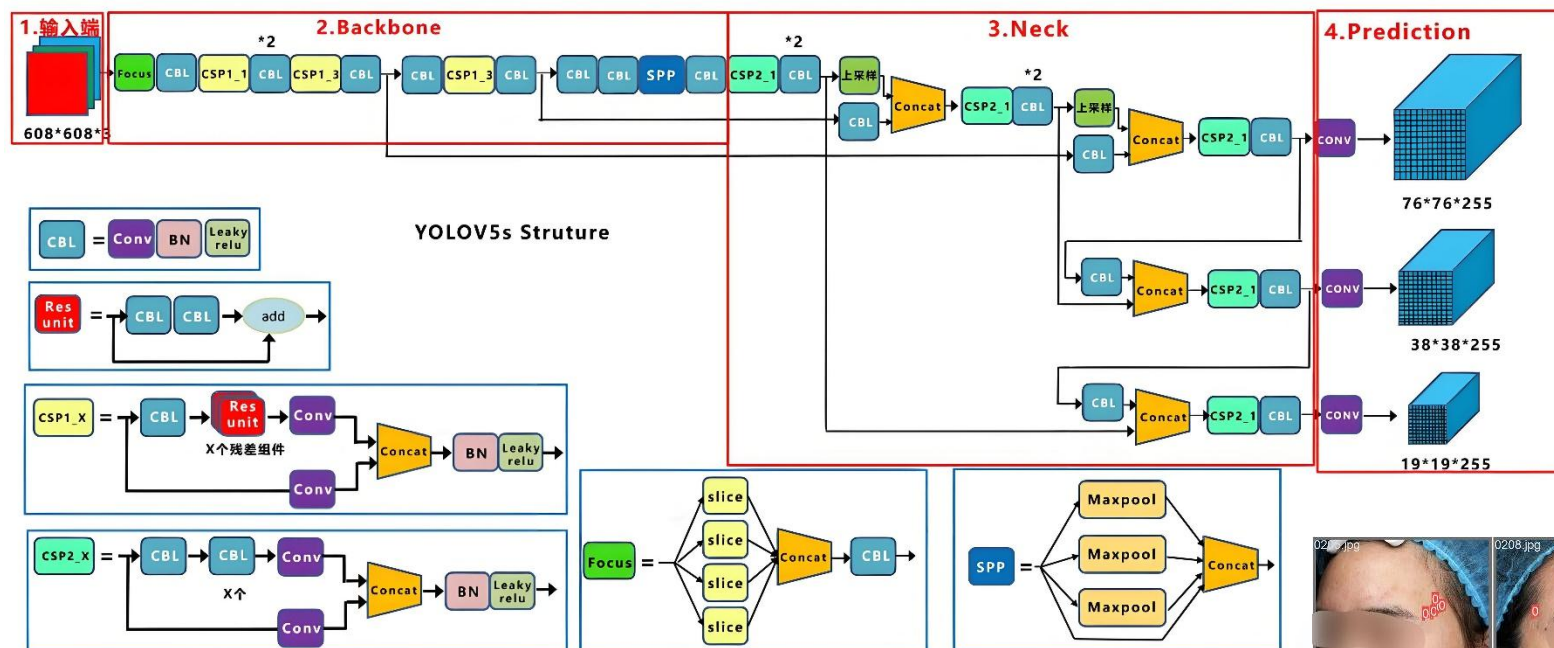
Why didn't Yolo perform very well?

➢ The target features are not obvious.

➢ Dense distribution leads to confusion.

➢ Complex background.

Part 3. Dataset Development

➢ Training and Validation Dataset: ACNE04 (80%)

➢ Test Dataset: ACNE04 (20%)

➢ 5-fold Cross-validation

Name	images	size	classes(including background)
ACNE04	1457	3112×3456	5
CelebAMask-HQ	30000	512×512	4
Flickr-Faces-HQ	1572	224×224	-
Nestlé Skin Health	4700	224×224	5
ACNE-Shanghai	322(309 selected)	3456×5184	5

Tab.1 information about 5 acne datasets

Class : Mild Number : 4 Class : Moderate Number : 10 Class : Severe Number : 24 Class : Very severe Number : 58

Class : Mild Number : 2 Class : Moderate Number : 7 Class : Severe Number : 29 Class : Very severe Number : 61
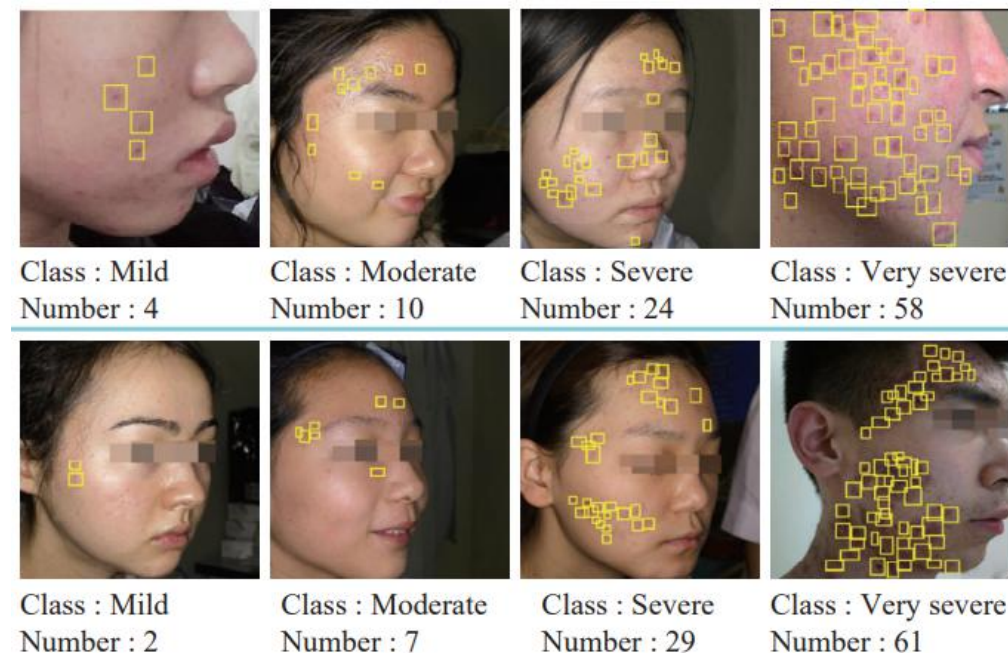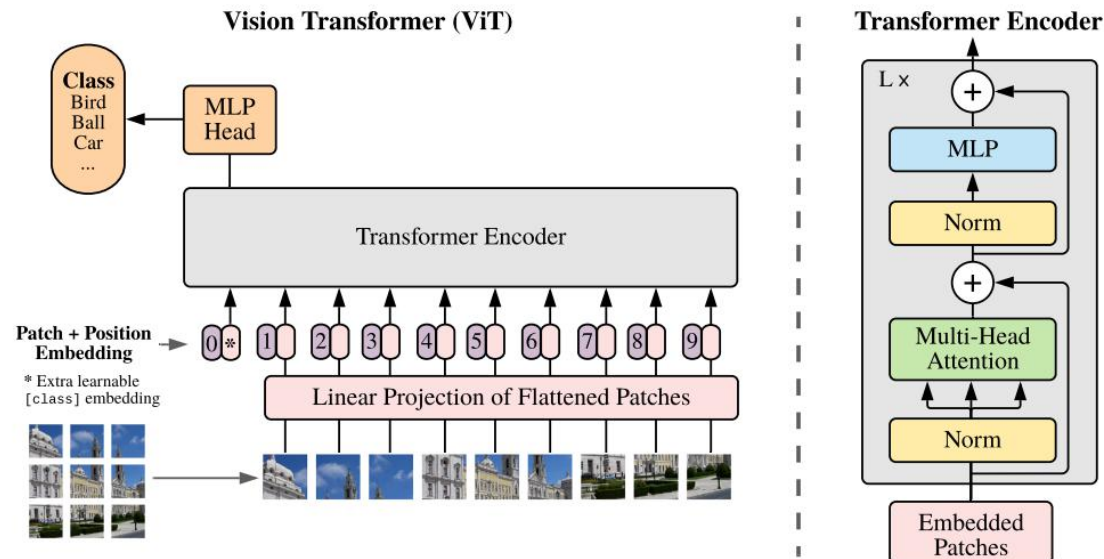
Fig. 4. Examples in the ACNE04 dataset. The numbers undereach image denote the ground-truth severity and lesion number.

Part 4.1 Model Selection : ViT

Vision Transformer (ViT)

An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale (ICLR 2021)

Model overview: We split an image into fixed-size patches, linearly embed each of them, add position embeddings, and feed the resulting sequence of vectors to a standard Transformer encoder

- **Advantages**
Global relationship modeling can enlarge the receptive field of image and obtain more context information
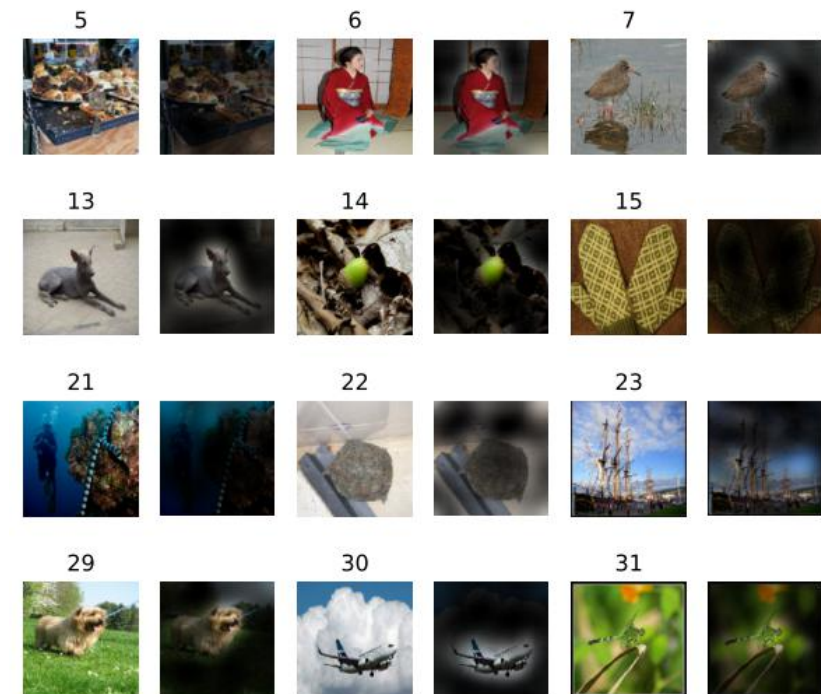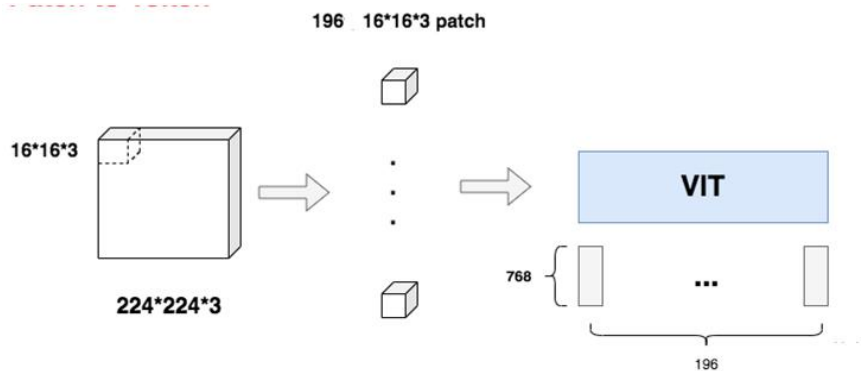

Fig.5 Attention maps

Part 4.1 Model Selection : ViT

- **Stage 1: Embedding**

  ➢ Patch Embedding $\quad x_p E \in \mathbb{R}^{N \times D}$

196   16*16*3 patch

16*16*3

VIT

224*224*3

768

196

  ➢ Class Embedding (Learnable Embedding) x_{cls}

  ➢ Position Embedding $\quad \mathbf{E}_{pos} \in \mathbb{R}^{(N+1) \times D}$

Final Input:

$$x_0 = [x_{cls}; x_p^1 E; x_p^2 E; \dots ; x_p^N E] + E_{pos}, \quad x_0 \in \mathbb{R}^{(N+1) \times D}$$

- **Stage 2: Transformer Encoder**
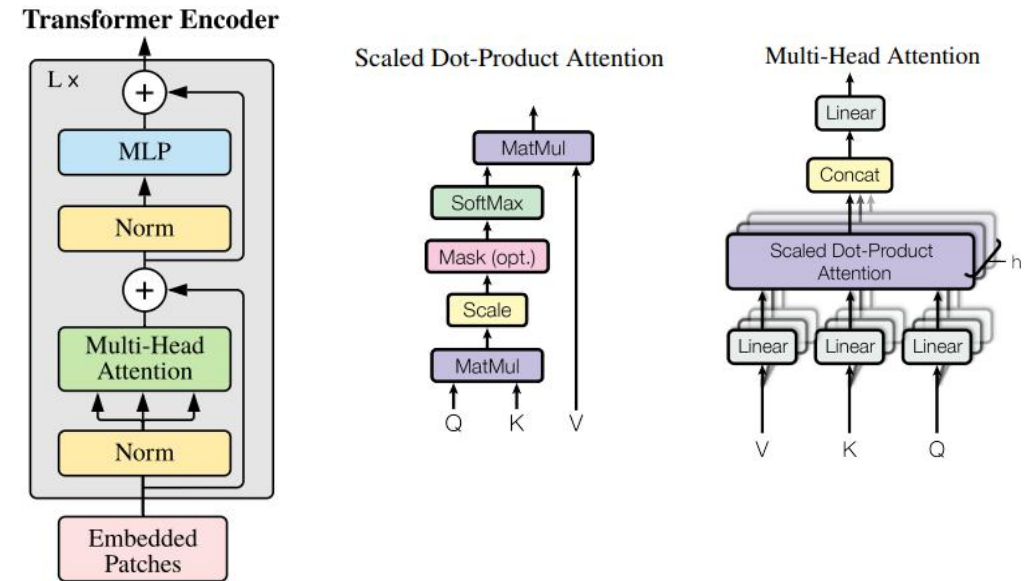
The Transformer encoder consists of alternating multi-head self-attention layers (MSA) and multi-layer perceptron blocks (MLP). Apply Layer Norm before each block and Residual Connection after each block

$$h(\mathbf{x}) = \underbrace{\mathbf{x}}_{\text{恒等函数}} + \underbrace{(h(\mathbf{x}) - \mathbf{x})}_{\text{残差函数}}$$
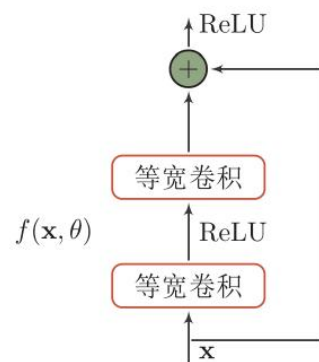
$f(\mathbf{x}, \theta)$

ReLU

等宽卷积

$f(\mathbf{x}, \theta)$ ReLU

等宽卷积

$\mathbf{x}$

Deep Residual Learning for Image Recognition
ILSVRC & COCO 2015 (nndl-book)

layer name	output size	18-layer	34-layer	50-layer	101-layer	152-layer
conv1	112×112	7×7, 64, stride 2				
		3×3 max pool, stride 2				
conv2_x	56×56	$\begin{bmatrix} 3\times3, 64 \\ 3\times3, 64 \end{bmatrix} \times 2$	$\begin{bmatrix} 3\times3, 64 \\ 3\times3, 64 \end{bmatrix} \times 3$	$\begin{bmatrix} 1\times1, 64 \\ 3\times3, 64 \\ 1\times1, 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1\times1, 64 \\ 3\times3, 64 \\ 1\times1, 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1\times1, 64 \\ 3\times3, 64 \\ 1\times1, 256 \end{bmatrix} \times 3$
conv3_x	28×28	$\begin{bmatrix} 3\times3, 128 \\ 3\times3, 128 \end{bmatrix} \times 2$	$\begin{bmatrix} 3\times3, 128 \\ 3\times3, 128 \end{bmatrix} \times 4$	$\begin{bmatrix} 1\times1, 128 \\ 3\times3, 128 \\ 1\times1, 512 \end{bmatrix} \times 4$	$\begin{bmatrix} 1\times1, 128 \\ 3\times3, 128 \\ 1\times1, 512 \end{bmatrix} \times 4$	$\begin{bmatrix} 1\times1, 128 \\ 3\times3, 128 \\ 1\times1, 512 \end{bmatrix} \times 8$
conv4_x	14×14	$\begin{bmatrix} 3\times3, 256 \\ 3\times3, 256 \end{bmatrix} \times 2$	$\begin{bmatrix} 3\times3, 256 \\ 3\times3, 256 \end{bmatrix} \times 6$	$\begin{bmatrix} 1\times1, 256 \\ 3\times3, 256 \\ 1\times1, 1024 \end{bmatrix} \times 6$	$\begin{bmatrix} 1\times1, 256 \\ 3\times3, 256 \\ 1\times1, 1024 \end{bmatrix} \times 23$	$\begin{bmatrix} 1\times1, 256 \\ 3\times3, 256 \\ 1\times1, 1024 \end{bmatrix} \times 36$
conv5_x	7×7	$\begin{bmatrix} 3\times3, 512 \\ 3\times3, 512 \end{bmatrix} \times 2$	$\begin{bmatrix} 3\times3, 512 \\ 3\times3, 512 \end{bmatrix} \times 3$	$\begin{bmatrix} 1\times1, 512 \\ 3\times3, 512 \\ 1\times1, 2048 \end{bmatrix} \times 3$	$\begin{bmatrix} 1\times1, 512 \\ 3\times3, 512 \\ 1\times1, 2048 \end{bmatrix} \times 3$	$\begin{bmatrix} 1\times1, 512 \\ 3\times3, 512 \\ 1\times1, 2048 \end{bmatrix} \times 3$
	1×1	average pool, 1000-d fc, softmax				
FLOPs		1.8×10^9	3.6×10^9	3.8×10^9	7.6×10^9	11.3×10^9

Fig.6 Framework of ResNet layers

Advantages:

➤ avoid loss in information transmission.

➤ transmit the input information to the output to protect the integrity of the information.

➤ simplify the learning goal and difficulty.

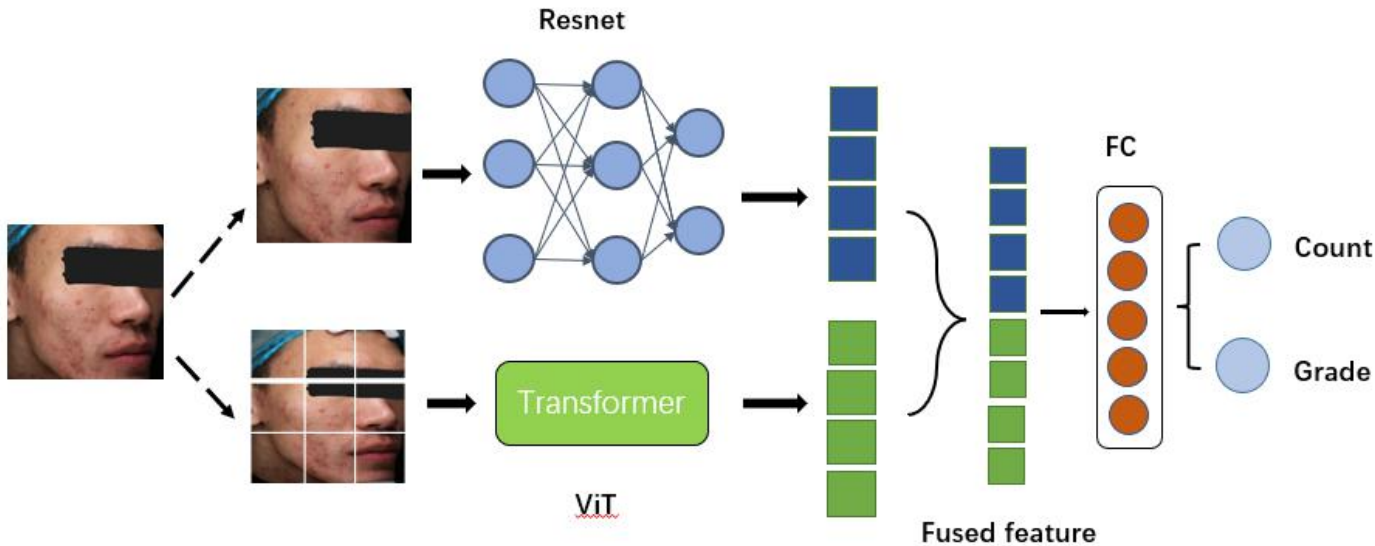# Part 4.3 Our Idea : FusedModel (ViT+Resnet)

Fig. 7 Model proposed by us. Using the idea of feature fusion, the features extracted from Resnet and Vit models are spliced and then passed through the fully connected layer

Advantages:

➢ Resnet allows the network to train deeper and learn more complex expressions.

➢ ViT is good at processing global information

Some Possible Shortcomings:

➢ The mode and degree of feature fusion is somewhat simple.

➢ The training curve is not very smooth.

Part 4.4 Loss Function and Optimization

- Firstly, we choose L2-norm as loss function:

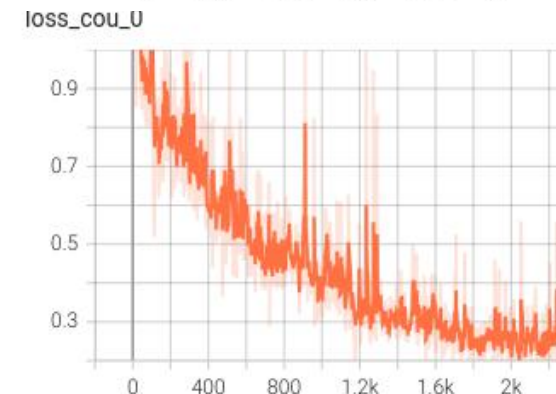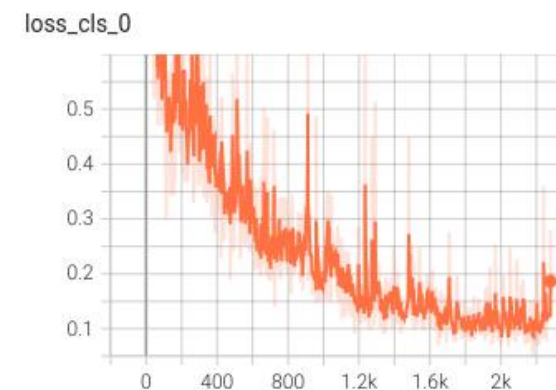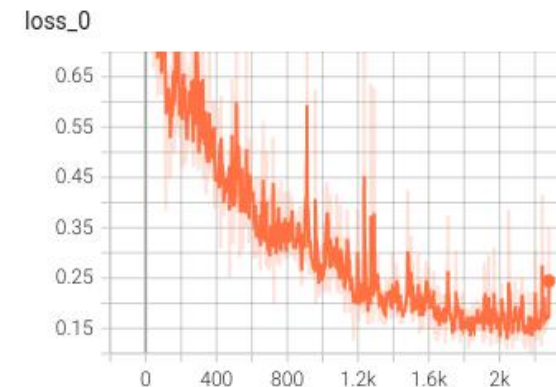$$L_{cou} = \sum_{i=1}^{n} (x_{cou}^{(i)} - y_{cou}^{(i)})^2$$

- **Add Kullback-Leibler Divergence:**

$$D_{KL}(P \,||\, Q) = \sum_{x} P(x) \, log \, \frac{P(x)}{Q(x)}$$

$$L_{cls} = \sum_{i=1}^{n} D_{KL}(x_{cou}^{(i)} - y_{cou}^{(i)})$$

$$Loss = \alpha \, L_{cou} + (1 - \alpha) \, L_{cls}$$

Optimizer: Adam

loss_0

loss_cls_0

loss_cou_U

Part 5. Experiment Result

	Precision	SE	SP	YI	MAE	MSE
class0	0.7863	0.8932	0.8677	0.7609	1.0680	1.6010
class1	0.8000	0.7874	0.8485	0.6359	2.8425	3.6121
class2	0.7692	0.5556	0.9766	0.5321	9.1111	13.7477
class3	0.8750	0.8077	0.9887	0.7964	7.0769	12.8333
avg / total	0.8076	0.7610	0.9204	0.6813	3.3664	6.6742

Tab.2 Prediction of FusedModel (KL-Loss) on acne counting

	Precision	SE	SP	YI		Name	Value
class0	0.8108	0.8738	0.8889	0.7627		learning rate	1e-4
class1	0.7923	0.8110	0.8364	0.6474		num of epoch	120
class2	0.7407	0.5556	0.9727	0.5282		batch size	64
class3	0.8750	0.8077	0.9887	0.7964		sigma	3
avg / total	0.8047	0.7620	0.9217	0.6837		alpha	0.6

Tab.3 Prediction of FusedModel (KL-Loss) on acne grading   Tab.4 HyperParameter Setting

Part 5. Comparison of Results

	Counting		Grading	
	AVG_ACC	**AVG_PRE**	**AVG_ACC**	**AVG_PRE**
ViT	0.5342	0.5090	0.5137	0.5342
ResNet	0.6615	0.6678	0.6712	0.6680
FusedModel (Ours)	**0.7979**	**0.8076**	**0.8014**	**0.8047**

Tab.5 Comparison of experimental results of different models

Conclusion:

➢ The experiment proves that our method is effective.

➢ In the future, we will try to integrate more abundant ways to make the advantages of the two models play a greater role

Part 6. References

- [CCID Dataset] Quattrini A, Boër C, Leidi T, Paydar R. A Deep Learning-Based Facial Acne Classification System. Clin Cosmet Investig Dermatol. 2022;15:851-857 https://doi.org/10.2147/CCID.S360450

- [ACNE04 Dataset] Wu, Xiaoping, Ni, Wen, Jie, Liang, Lai, Yu-Kun, Cheng, Dongyu, She, Ming-Ming, & Yang, Jufeng. (2019). Joint Acne Image Grading and Counting via Label Distribution Learning. In *IEEE International Conference on Computer Vision*.

- [Nestlé Skin Health Dataset] https://github.com/microsoft/nestle-acne-assessment

- [Flickr-Faces-HQ Dataset] https://github.com/HuynhThanhQuan/skin-detective

Part 6. References

- Islam, Md Baharul, Masum Shah Junayed, Arezoo Sadeghzadeh, Nipa Anjum, Afsana Ahsan, A. F. M. Shahen Shah. (2023). Acne Vulgaris Detection and Classification: A Dual Integrated Deep CNN Model.

- Lin, Yi, Jingchi Jiang, Dongxin Chen, Zhaoyang Ma, Yi Guan, Xiguang Liu, Haiyan You和 Jing Yang. DED: Diagnostic Evidence Distillation for Acne Severity Grading on Face Images. Expert Systems with Applications 228 (2023): 120312.

Thanks