PROJECT PRESENTATION

Presented by: Nay Win Hlaing, Khin Su Myat Moe
ConvNext Group

Transformers see long-range relationships in images

Transformer-inspired

Redesign of ResNet ideas

USE CASES

- High-accuracy classification tasks
- Fine-grained recognition
- Image captioning (as backbone)
- Transfer learning in modern applications

FIELDS

- Autonomous vehicles
- AI startups
- Computer vision R&D
- Next-gen mobile vision apps

CONVNEXT VARIANTS

- ConvNeXt-Tiny
- ConvNeXt-Base
- ConvNeXt-Small
- ConvNeXt-Large
- ConvNeXt-XL

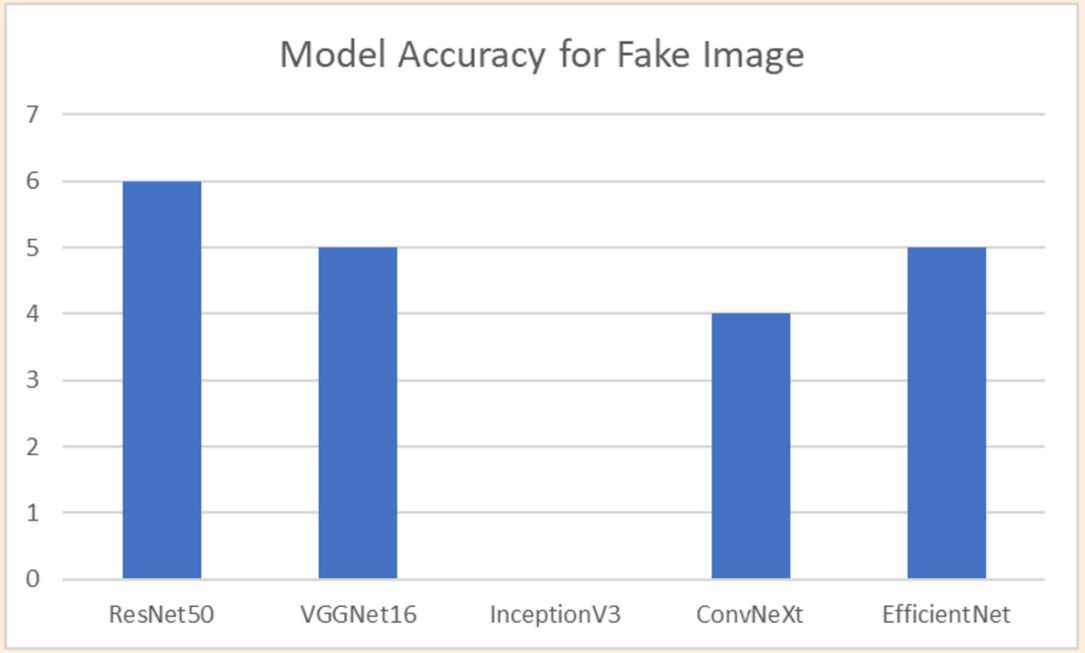
ARCHITECTURE

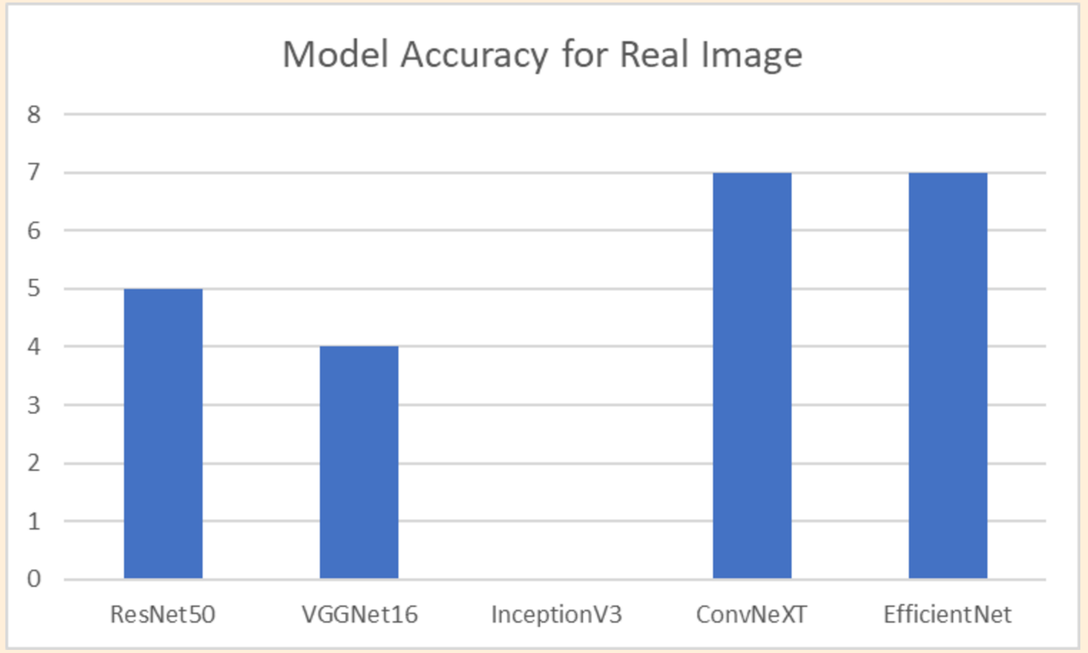
- 4 stages (ResNet)
- Total 18 blocks - [3,3,9,3]
- Helps learn features gradually at different scales

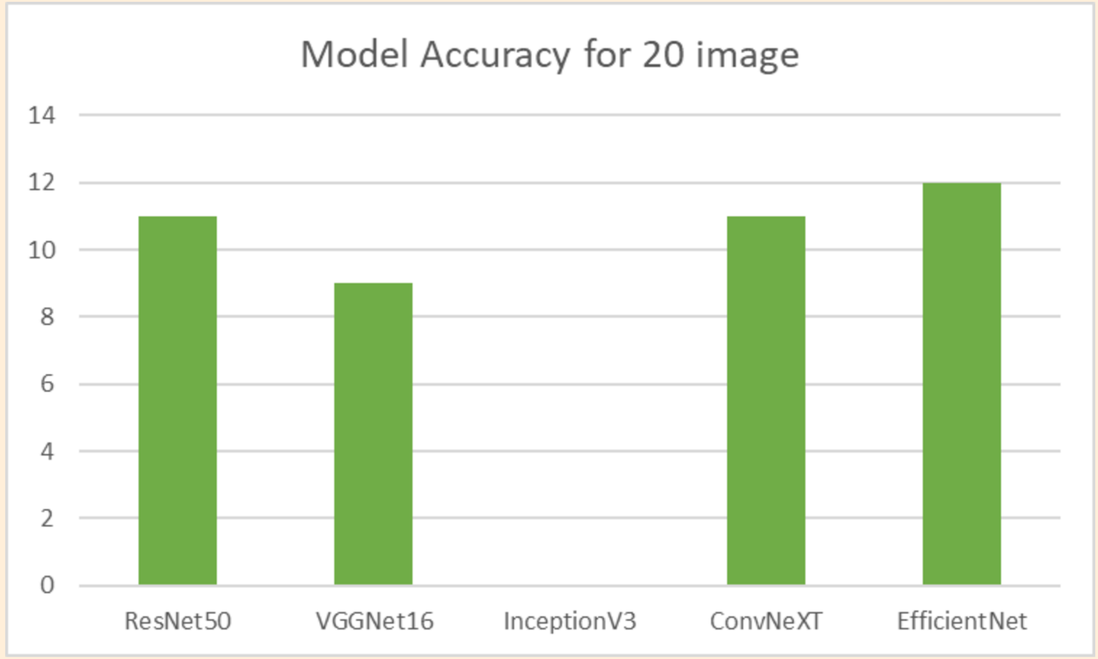
Each block :

- Kernel - 7 x 7
- GELU Activation
- Residual connection (ResNet)
- Layer Normalization (Transformers)
  - Depends on features instead of batch size
  - More accurate, stable, faster in training
- Depthwise Convolutions
  - 1 filter per input channel
  - Keeps spatial info
- Fewer downsampling steps
  - Preserves more details and spatial info

Step	Input Shape	Output Shape	Notes
Stem	224×224×3	56×56×96	4×4 Conv with stride 4
Stage 1	56×56×96	56×56×96	3 ConvNeXt blocks
Downsample	56×56×96	28×28×192	stride-2 Conv
Stage 2	28×28×192	28×28×192	3 blocks
Downsample	28×28×192	14×14×384	
Stage 3	14×14×384	14×14×384	9 blocks
Downsample	14×14×384	7×7×768	
Stage 4	7×7×768	7×7×768	3 blocks
Global Pool	7×7×768	768	Spatial average
Dense	768	0	Classification output


Model Accuracy for Fake Image
7
6
5
4
3
2
1
0
ResNet50
VGGNet16
InceptionV3
ConvNeXt
EfficientNet


Model Accuracy for Real Image
8
7
6
5
4
3
2
1
0
ResNet50
VGGNet16
InceptionV3
ConvNeXT
EfficientNet

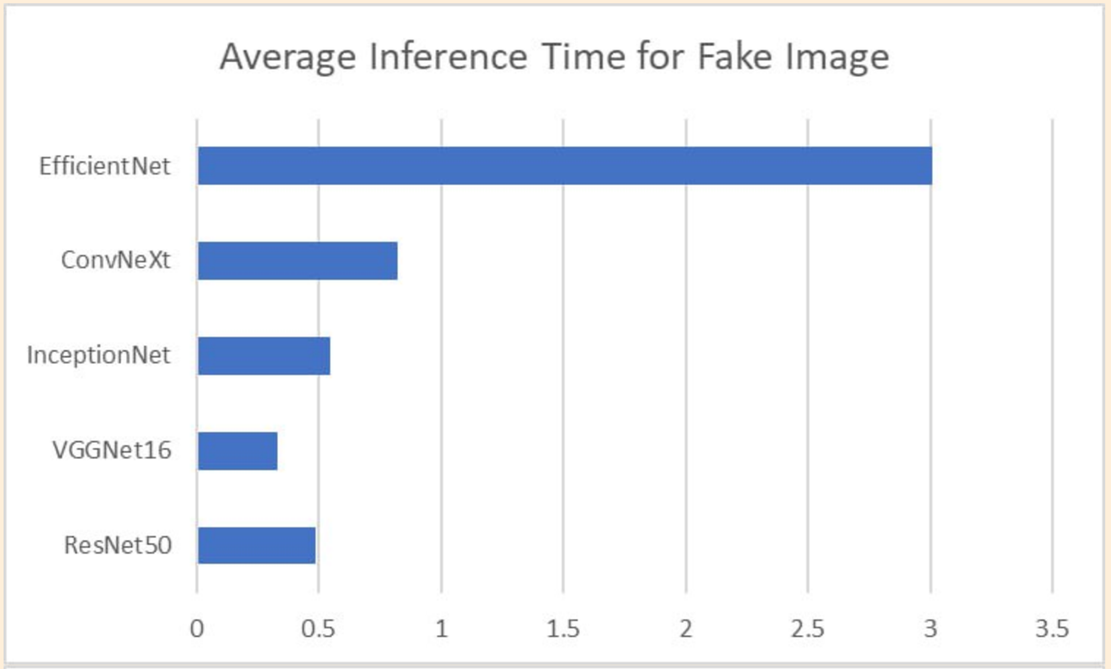

Model Accuracy for 20 image
14
12
10
8
6
4
2
0
ResNet50
VGGNet16
InceptionV3
ConvNeXT
EfficientNet

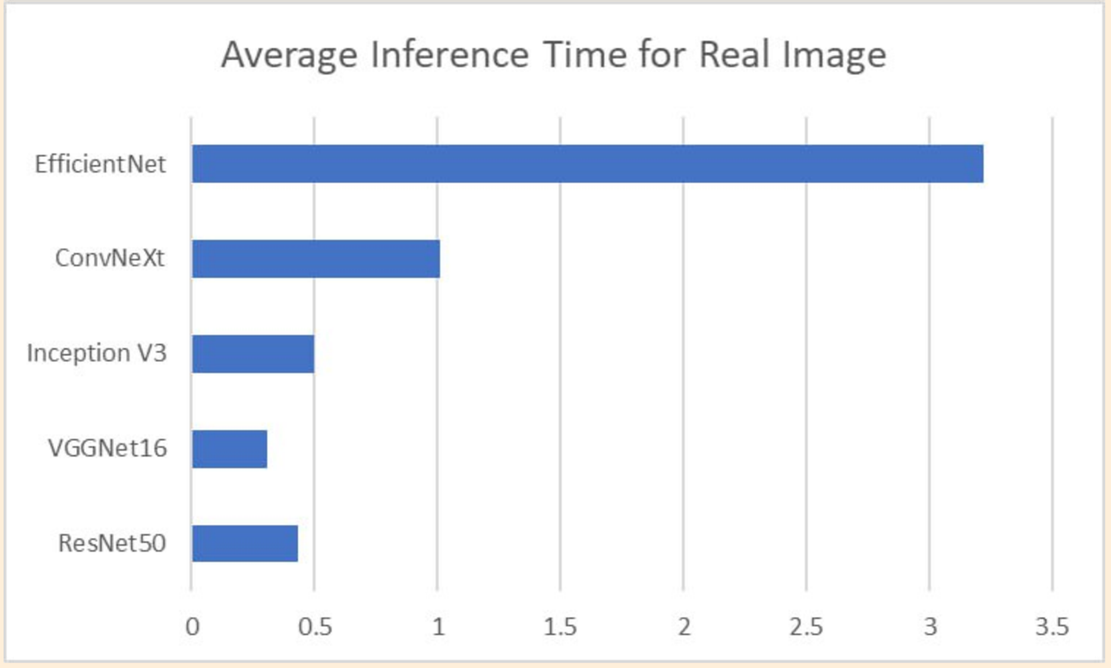
ResNet50	VGGNet16	InceptionV3	ConvNeXt	EfficientNet	label
basset	basset	web_site	basset	basset	Basset
American_Staffordsh	boxer	stopwatch	bull_mastiff	bull_mastiff	Bulldog
Chihuahua	Chihuahua	web_site	Chihuahua	Chihuahua	Chihuahua
Pembroke	Pembroke	web_site	Pembroke	Pembroke	Corgi
Doberman	Doberman	hammer	Weimaraner	dalmatian	Dachshund
German_shepherd	German_shepherd	flatworm	Norwegian_elkhour	schipperke	German Shepherd
Eskimo_dog	Eskimo_dog	flatworm	Eskimo_dog	Eskimo_dog	Husky
Labrador_retriever	Rhodesian_ridgebac	web_site	Weimaraner	Labrador_retrieve	Labrador
Maltese_dog	Maltese_dog	web_site	Maltese_dog	Maltese_dog	Maltese
basenji	dingo	web_site	Eskimo_dog	Eskimo_dog	Shiba Inu

Synthetic Picture Prediction

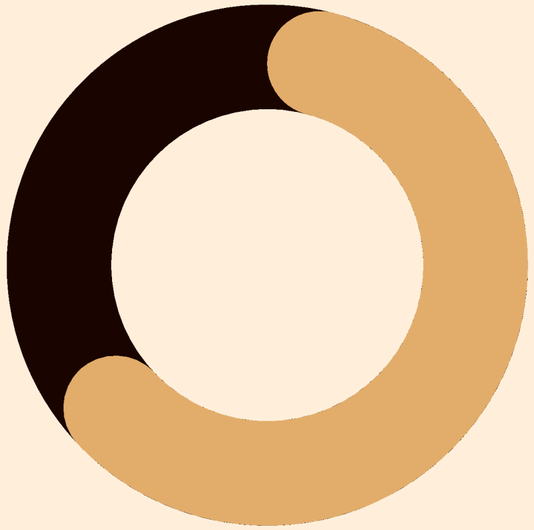
ResNet50	VGGNet16	InceptionV3	ConvNeXt	EfficientNet	label
basset	basset	clog	bloodhound	basset	Basset Hound
Saint_Bernard	Great_Dane	pitcher	French_bulldog	French_bulldog	Bulldog
Chihuahua	Chihuahua	web_site	Chihuahua	Chihuahua	Chihuahua
Pembroke	Pembroke	web_site	Pembroke	Pembroke	Corgi
black-and-tan_coonhound	black-and-tan_coonhound	web_site	Weimaraner	Labrador_retriever	Dachshund
German_shepherd	German_shepherd	iron	German_shepherd	Saluki	German Shepherd
Eskimo_dog	Eskimo_dog	hammer	Siberian_husky	Siberian_husky	Husky
Labrador_retriever	Great_Dane	pitcher	Labrador_retriever	Labrador_retriever	Labrador
Norfolk_terrier	Lhasa	pitcher	Maltese_dog	Maltese_dog	Maltese
basenji	dingo	web_site	Arctic_fox	schipperke	Shiba Inu

Real Picture Prediction


Average Inference Time for Fake Image
EfficientNet
ConvNeXt
InceptionNet
VGGNet16
ResNet50
0
0.5
1
1.5
2
2.5
3
3.5


Average Inference Time for Real Image
EfficientNet
ConvNeXt
Inception V3
VGGNet16
ResNet50
0
0.5
1
1.5
2
2.5
3
3.5

OBESRVATION OF TOP 3 PREDICTIONS



- Shiba Inu & Dachshund: Neither synthetic nor real images ranked in top-3 for any model.
- German Shepherd: EfficientNet failed to include it in top-3 for both image types.
- When the correct label is NOT the top prediction - Synthetic Images Show Higher Top-3 Consistency Than Real Images

CONCLUSION

EfficientNet - Lowest Speed with Highest Accuracy
VGGNet - Highest Speed with 2nd Lowest Accuracy
Inception - No Accuracy

Covnext performs better on real images - Training data is based on real world images.

THANK YOU