Evaluating GAN-Based Augmentation Across Deep Learning Architectures for Endoscopic Image Classification

Laimonas Janutėnas Dmitrij Šešok
Vilnius Gediminas Technical University, Vilnius, Lithuania

Summary: The benefit of synthetic training data varies across classifiers. We analyse published Kvasir v2 results for five CNNs, comparing classification accuracy, reported gain and relative reduction in classification error. Enhanced StarGAN adds synthetic training images. All five models improve, with gains of **0.298–0.704 percentage points**. VGG-16 showed the largest gain, while EfficientNet-B7 achieved the highest final accuracy (**95.25%**).

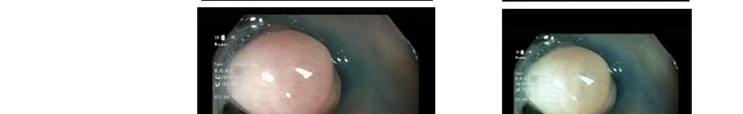
ENDOSCOPIC DATA & KVASIR V2

Endoscopic images show the gastrointestinal tract's inner surface. Tissue colour, texture and anatomy support classification, while viewpoint and illumination vary. Kvasir v2 is a public dataset of landmarks, pathological findings and polyp-removal scenes. We compare five CNNs to study the benefit of synthetic training data.

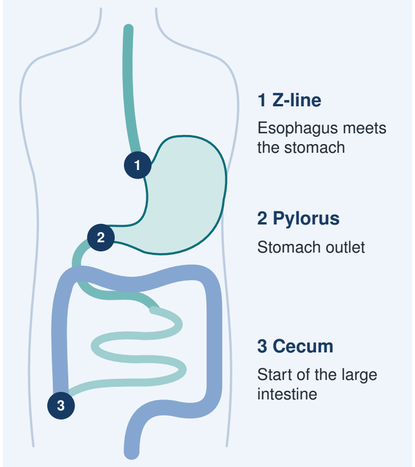
8,000	8	1,000
real JPEG images	classes	images per class

Landmarks: normal cecum, normal pylorus, normal Z-line.
Pathological findings: esophagitis, polyps, ulcerative colitis.
Polyp-removal scenes: dyed lifted polyps, dyed resection margins.

Augmented training set: 11,200 images



Original JPEG images range from 720×576 to 1920×1072 pixels.



SYNTHETIC IMAGE AUGMENTATION

Enhanced StarGAN generates additional samples with varied viewing geometry. The Perspective Transformation Module learns spatial warps; the Viewpoint Attention Module emphasizes informative feature locations. These act between encoding and residual processing. Synthetic images expand the training set.



Figure 1. Enhanced StarGAN generator. Teal blocks add spatial transformation and attention.

Figure 2. Original (left), StarGAN (middle) and enhanced model (right).

RELATIVE ERROR REDUCTION

This measure expresses the decrease in error as a share of the reference error rate. For EfficientNet-B7, error falls from 5.38% to 4.75%, a relative reduction of about **11.7%**.

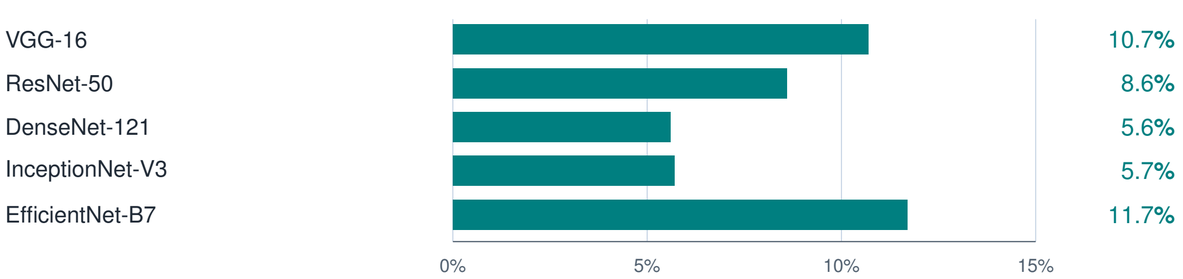


Figure 5. Approximate error reduction, computed by dividing the accuracy increase by the reference error rate.

CLASSIFICATION ACCURACY

Accuracy measures correct classification across all eight classes. Each pair compares a published reference score with the corresponding enhanced-augmentation result, showing final performance alongside the gain from synthetic training data.

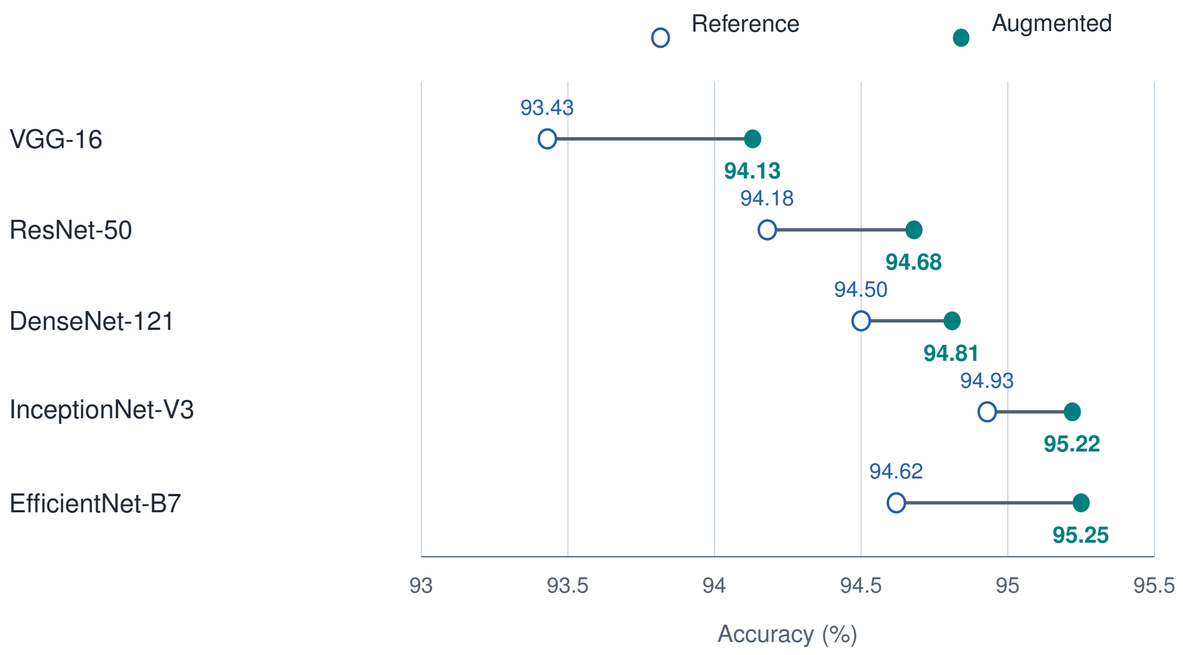


Figure 3. Reported reference and augmented accuracies.

Table 1. Full accuracy results for all five classifiers.

Classifier	Reference (%)	Augmented (%)	Gain (pp)
VGG-16	93.43	**94.13**	+0.704 ± 0.118
ResNet-50	94.18	**94.68**	+0.504 ± 0.106
DenseNet-121	94.50	**94.81**	+0.318 ± 0.131
InceptionNet-V3	94.93	**95.22**	+0.298 ± 0.082
EfficientNet-B7	94.62	**95.25**	+0.636 ± 0.097

ACCURACY GAIN & UNCERTAINTY

All five architectures show positive reported gains. Improvements range from 0.298 to 0.704 percentage points. The whiskers show the reported uncertainty around each gain.

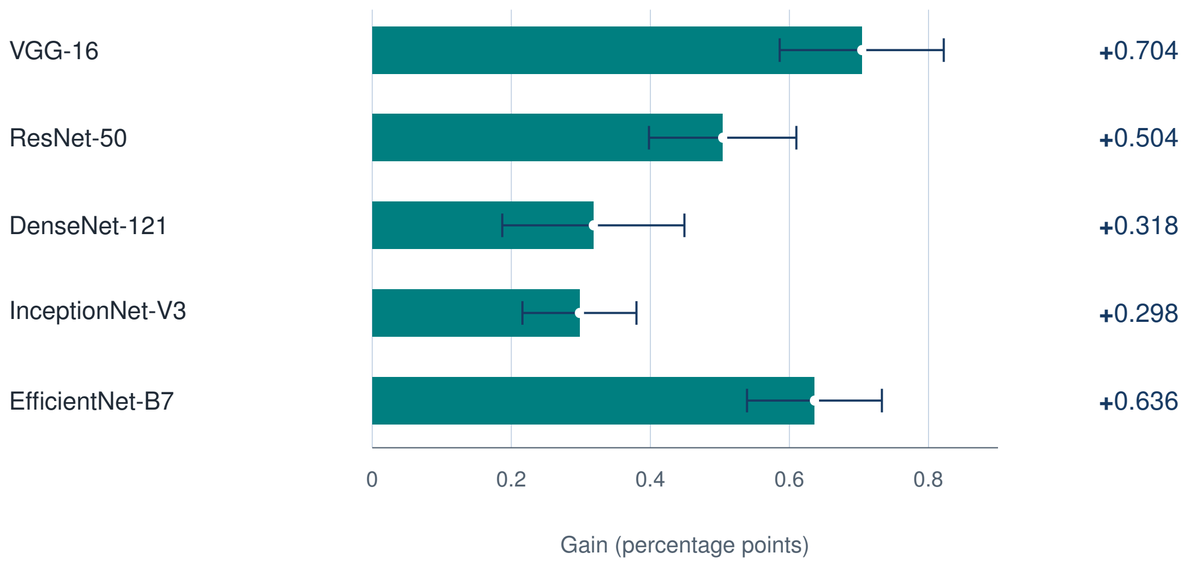


Figure 4. Whiskers show the published confidence intervals (confidence level unspecified). Gains are shown as reported.

CONCLUSIONS

All five classifiers achieved higher accuracy with enhanced StarGAN augmentation than their published reference scores. VGG-16 shows the largest absolute gain (**+0.704 pp**). EfficientNet-B7 combines the highest accuracy (**95.25%**) with the largest relative error reduction (about **11.7%**). When comparing classifiers, consider both the reported accuracy gain and the accuracy achieved with augmentation.