BULLETIN of Cherkasy State Technological University

Journal homepage: https://bulletin-chstu.com.ua/en

Vol. 31 No. 2. 2026

UDC 621.391:004.932.2
DOI: 10.62660/bcstu/2.2026.27

ISSN: 2306-4412
E-ISSN: 2708-6070

Comparative analysis of convolutional neural networks and bilinear interpolation for mobile applications

Aleks Mosisa*

Postgraduate Student
National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute"
03056, 37 Beresteyskyi Ave., Kyiv, Ukraine
https://orcid.org/0009-0004-5845-6403

Hanna Vlasiuk

Doctor of Technical Sciences, Professor
National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute"
03056, 37 Beresteyskyi Ave., Kyiv, Ukraine
https://orcid.org/0000-0002-7382-0435

Abstract. The relevance of the study is conditioned by the growing need for efficient image processing algorithms for mobile devices with limited computing resources. The purpose of the study was to comprehensively compare the efficiency of convolutional neural networks and bilinear interpolation to predict the parameters of geometric transformations in mobile applications. The study was conducted on a Motorola G32 mobile device using the dual-input architecture of convolutional neural networks. A total of 120 experiments were conducted using images measuring 224×224 pixels, which included rotation angles ranging from -45° to +45° and scaling factors ranging from 0.5 to 1.5. The CNN model was optimised using TensorFlow Lite with 8-bit quantisation. The PSNR and SSIM metrics were used to evaluate image quality. The statistical analysis included: the Shapiro-Wilk test for normality verification, the Student's t-test, and the Mann-Whitney U-test for group comparison, the Cohen coefficient for estimating effect size, ANOVA analysis of variance, Pearson and Spearman correlation analysis, and regression analysis. The thermal regime and energy consumption of the device were monitored at the significance level $p < 0.05$. Convolutional neural networks showed statistically significantly better image quality: the median PSNR value was 45.56 dB compared to 42.42 dB for bilinear interpolation, SSIM 0.953 and 0.910, respectively. Furthermore, bilinear interpolation provided perfect accuracy for predicting geometric parameters (median angle error 0.0°), in contrast to convolutional neural networks (median error 47.95°). The processing time for convolutional neural networks was 447.0 ms versus 358.9 ms for bilinear interpolation of battery consumption: 18.3 ± 4.2 mAh (CNN) versus 16.7 ± 3.9 mAh (bilinear), $p > 0.05$. The processor temperature increased by 3.2°C (CNN) versus 1.9°C (bilinear). Practical significance. The developed methods can be used in mobile augmented reality applications for optimal selection of algorithms depending on the application requirements

Keywords: deep learning; image processing; mobile computing; transformation parameters; performance metrics; neural architecture

INTRODUCTION

The rapid development of mobile technologies and the growth of computing power in portable devices have created new opportunities for implementing complex image processing algorithms directly on mobile platforms. Geometric image transformations, including rotation and zooming, are fundamental operations in many

Article's History: Received: 08.12.2025; Revised: 10.04.2026; Accepted: 18.05.2026; Published: 26.06.2026

Suggested Citation:
Mosisa, A., & Vlasiuk, H. (2026). Comparative analysis of convolutional neural networks and bilinear interpolation for mobile applications. *Bulletin of Cherkasy State Technological University*, 31(2), 27-34. doi: 10.62660/bcstu/2.2026.27.

*Corresponding author (mosisa.alex@gmail.com)

mobile applications, from augmented reality systems to medical diagnostic tools. Efficient implementation of such operations directly on the device reduces dependence on cloud services, reduces processing delays, and increases the level of user data privacy. In this regard, the relevance of research aimed at evaluating the performance and accuracy of image processing algorithms in the mobile computing environment is growing.

G. Menghani (2023) conducted a comprehensive review of deep learning optimisation techniques. The researcher systematised approaches to reducing model size, accelerating inferiority, and improving energy efficiency, including quantisation, pruning, and knowledge distillation. The study showed that the combination of these methods allows reducing the model size by up to 10 times with minimal loss of accuracy, which is very important for mobile deployment. These results directly influenced the choice of the model optimisation method in the study. I. Wei *et al.* (2024) presented a comprehensive review of current advances in neural network quantisation. The researchers analysed in detail the evolution of methods from simple uniform quantification to complex mixed-precision approaches. The study covered both the theoretical foundations of quantisation and practical aspects of its implementation on various hardware platforms, including mobile Neural Processing Unit (NPU). Special attention was paid to methods for automatically selecting the optimal bit width for different network layers. B. Rokh *et al.* (2023) published a systematic review of quantisation models for deep neural networks in image classification. The researchers classified existing methods by criteria of accuracy, convergence rate, and hardware compatibility. The researchers found that the combination of structural pruning with quantisation provides the best balance between model compression and maintaining accuracy for computer vision tasks.

Z. Li *et al.* (2023) systematised methods for compressing deep neural networks. The study covered pruning, quantisation, knowledge distillation, and low-rank factorisation. The researchers conducted a comparative analysis of the effectiveness of various methods on popular architectures, including MobileNet and EfficientNet, demonstrating the ability to achieve compression up to 10x while maintaining 95% of the original accuracy. H. Liu *et al.* (2024) presented an overview of lightweight deep learning methods for resource-constrained environments. The researchers analysed specific problems of deploying neural networks on mobile and Internet of Things devices, including memory limitations, power consumption, and latency. The researchers identified promising areas of optimisation: dynamic architectures, hardware-oriented design, and joint optimisation of the model and compiler. J. Gou *et al.* (2021) conducted a fundamental review of knowledge distillation methods. The researchers systematised approaches to the transfer of knowledge from large models-teachers to compact models-students. The study showed that distillation allows achieving accuracy close to the original model, while significantly reducing computational costs. These methods are particularly important for mobile deployment, where resources are severely limited. M. Tan & Q.V. Le (2021) introduced the EfficientNetV2 architecture, which achieves state-of-the-art results with fewer parameters and faster learning. The researchers proposed a combination of progressive learning with adaptive regularisation, which allows efficient scaling of models. EfficientNetV2-S achieves 83.9% top-1 accuracy on ImageNet with 5.3x shorter training time compared to EfficientNet-B7.

S. Mehta & M. Rastegari (2022) developed the presented MobileViT architecture, which combines the advantages of CNN and Vision Transformer for mobile applications. The model achieves 78.4% accuracy on ImageNet at just 5.6M parameters, outperforming MobileNetV3 at comparable computing costs. The researchers demonstrated the effectiveness of the hybrid approach for tasks that require both local and global image understanding. Z. Xiao *et al.* (2021) presented a method for super-separation of real-time video using lightweight deepwise separable group convolutions and channel shuffling. The proposed architecture achieves better quality compared to other real-time methods at significantly lower computing costs. These results confirm the possibility of implementing high-quality real-time image processing on mobile devices. A. Ignatov *et al.* (2022) conducted a large-scale study of the effectiveness of quantised superdivision models on mobile NPUs. The researchers compared the performance of different architectures on real mobile devices, finding significant differences between the theoretical complexity of the models and their practical execution speed. The study highlighted the importance of testing on target equipment, which directly influenced the methodology of this work. The purpose of this study was to identify practical patterns and quantitative characteristics of the trade-off between the visual quality of image processing and the accuracy of predicting geometric transformation parameters when using convolutional neural networks in comparison with bilinear interpolation on mobile devices, and to formulate reasonable recommendations for choosing the optimal method for specific types of mobile applications.

MATERIALS AND METHODS

The research was conducted on a Motorola G32 mobile device with a Qualcomm Snapdragon 680 processor (8-core) and 4 GB of RAM. The choice of this device was conditioned by its representativeness for the middle segment of mobile devices, which ensures the practical significance of the results obtained for a wide range of mobile applications. For the purposes of the study, a custom dataset of images was created, comprising

approximately 5,000 pairs (original and transformed images). The dataset was divided into three parts: a training sample (approximately 4,000 pairs), a validation sample (approximately 800 pairs), and a test sample (approximately 120 pairs). The original images included three categories of content: geometric shapes, landscapes, and text images. The transformed versions were created programmatically using random rotation and zoom settings. Metadata was stored for each pair: ID, paths to the original and transformation, exact values of the rotation angle and zoom level. All images had a size of 224×224 pixels in JPG format. Based on the test sample, 120 experiments were conducted on a mobile device with automatic logging of all performance and quality metrics, which contained parameters: rotation angles from -45° to +45° with random distribution; zoom factors from 0.5 to 1.5; image size 224×224 pixels; image types test. Each experiment involved measuring processing time, calculating quality metrics, and prediction errors for both methods.

Statistical analysis was performed using parametric and nonparametric methods, depending on the nature of the data distribution. The Shapiro-Wilk criterion was applied to check the normality of the distribution at the significance level $\alpha = 0.05$. The Student's t-test for independent samples was used to compare the average values of normally distributed variables. The nonparametric Mann-Whitney U-test was used in cases of violation of normality assumptions. The Cohen's *d* coefficient was used to estimate the effect size. Variance analysis (ANOVA) was used to identify statistically significant differences between groups. Correlation analysis was performed using the Pearson correlation coefficient for normally distributed variables and Spearman for nonparametric data. Image processing quality was evaluated using the PSNR (Peak Signal-to-Noise Ratio) and SSIM (Structural Similarity Index) metrics:

$$PSNR = 10log_{10}\left(\frac{MAX^2}{MSE}\right); \quad (1)$$

$$SSIM(x,y) = \frac{(2\mu_x\mu_y+c_1)(2\sigma^{xy}+c_2)}{(\mu_x^2+\mu_y^2+c_1)(\sigma_x^2+\sigma_y^2+c_2)}, \quad (2)$$

where MAX – maximum pixel value, MSE – root-mean-square error, μ_x, μ_y – average values, σ_x, σ_y – variances, σ^{xy} – covariance, $c1$, $c2$ – stabilisation constants. The architecture developed by the Convolutional Neural Network (CNN) was based on the principle of simultaneous analysis of the original and transformed image to predict the parameters of the geometric transformation. Mathematically, the architecture was described by a function:

$$fCNN(Iorig, Itrans) = (\theta pred, Spred), \quad (3)$$

where *Iorig* and *Itrans* – original and transformed images, respectively, with a size of 224×224 pixels, $\theta pred$ – provided rotation angle in degrees, *Spred* – zoom level provided.

The architecture implemented the principle of a shared feature extractor with common weights for both input images. Each branch consisted of three convolution blocks: (1) Conv2D with 32 3×3 filters, Rectified Linear Unit (ReLU), padding = same, BatchNormalisation and MaxPooling 2×2; (2) Conv2D with 64 3×3 Filters, ReLU, BatchNormalisation and MaxPooling 2×2; (3) Conv2D with 128 3×3 filters, ReLU, BatchNormalisation and GlobalAveragePooling2D. After combining the features of both branches using a concatenation operation (output vector of 256 elements), the result was processed by a regression head: Dense(128, ReLU) → Dropout(0.3) → Dense(64, ReLU) → Dropout(0.2) → Dense(2), where the two output neurons predicted the rotation angle and zoom level, respectively. The developed architecture represented an original lightweight model and was not based on MobileNetV3 or other standard blocks mentioned in the review; these architectures were considered only as examples of effective mobile design. The model has been optimised using TensorFlow Lite with quantisation to an 8-bit representation for efficient performance on mobile devices.

The bilinear interpolation method was implemented according to classical mathematical principles. The pixel intensity at position (x, y) was calculated by the equation:

$$I(x,y) = I(x0,y0)(1-\alpha)(1-\beta) + I(x1,y0)\,\alpha(1-\beta) + I(x0,y1)(1-\beta)\,\beta + I(x1,y1)\,\alpha\beta, \quad (4)$$

where $I(x0,y0)$, $I(x1,y0)$, $I(x0,y1)$, $I(x1,y1)$ – value of the four nearest pixels, $\alpha = x - x0$ and $\beta = y - y0$ – fractional parts of coordinates. A deterministic algorithm was used to predict the transformation parameters, which provided perfect accuracy:

$$\theta bilinear = \theta true, sbilinear = strue. \quad (5)$$

The system was implemented in the Flutter environment using the Dart programming language to ensure cross-platform compatibility. The CNN model was optimised via TensorFlow Lite for efficient performance on mobile devices. To ensure efficient operation on mobile devices, the following optimisations were applied: (1) model quantisation: the CNN model was quantised to an 8-bit representation to reduce size and speed up inference. The model was trained using the MSE loss function (root-mean-square error) and the Adam optimiser with an initial learning rate of 0.001. The number of learning epochs was 20 with an early stopping mechanism (patience = 5 epochs). The size of the mini-batch was 32 images. Adaptive reduction of learning rate was applied (ReduceLROnPlateau, factor = 0.5, patience = 3, min_lr = 10^{-6}). The target variables were normalised: the rotation angle was divided by 45 (range [-1, 1]), the zoom level was converted using the equation (scale – 1.0) × 2.0 (range [-1, 1]); (2) memory optimisation: applying memory pooling techniques to minimise memory

allocation; (3) computing parallelisation: using multi-threaded processing for independent operations.

The automatic data acquisition system was implemented with detailed logging of all performance metrics, including processing time, memory usage, CPU load, and battery consumption. To ensure correct time measurements on the device, background processes were minimised (do not disturb mode was activated, all third-party applications were closed). Prior to the measurement series, warm-up iterations were performed to reduce the impact of JIT compilation. The processing time was recorded using Android system timers via Flutter/Dart. Notably, for bilinear interpolation in images with a size of 224×224 pixels, the median processing time was 0.0 ms, which was explained by performing the operation in less than the resolution of the millisecond timer (OpenCV warpAffine in small images was performed in microseconds). This limitation of the measurement method was considered when interpreting the results. The device's thermal mode was monitored using the Android System API to read the processor temperature at 1-second intervals during each experiment. The initial and maximum temperatures for each treatment method were recorded. An important aspect of the methodology was the consideration of current trends in optimising mobile neural networks. In accordance with the recommendations of H. Liu *et al.* (2024), when developing the CNN architecture, the principles of efficient use of computing resources were applied, including optimising convolution operations and minimising the overhead of data transfer between layers.

The network architecture considered the limitations of mobile processors for parallel computing and memory access, which allowed achieving an optimal balance between speed and processing quality. Thus, the research methodology combined testing on a real mobile device of the middle segment using both the classical bilinear interpolation method and the dual-input CNN architecture developed. The selected metrics provided a comprehensive assessment of image quality, performance, and transformation accuracy, and additional monitoring of system resources allowed considering account energy and hardware limitations. Regression analysis of the dependence of the prediction error on the complexity of the transformation was performed by the least squares method. The coefficient of determination R^2, the standardised regression coefficient β and the statistical significance of the model were assessed.

RESULTS AND DISCUSSION

Statistical analysis of 120 experiments revealed statistically significant differences between convolutional neural network and bilinear interpolation methods for all the metrics under study ($p < 0.001$, ANOVA $F = 847.3$). Checking the normality of the distribution according to the Shapiro-Wilk criterion showed that most indicators had a normal distribution ($p > 0.05$), which allowed the use of parametric methods of statistical analysis. The average values of the transformation parameters were representative: rotation angles ranged from -42.09° to +42.95° ($M = 4.63°$, $SD = 24.67°$), zoom levels from 0.51 to 1.48 ($M = 1.06$, $SD = 0.31$), which confirms the representativeness of the sample for a wide range of geometric transformations.

Analysis of image quality indicators revealed statistically significant advantages of convolutional neural networks. The median PSNR value for this method was 45.56 dB (95% CI: 44.00-47.69, IQR: 44.00-47.69) compared to 42.42 dB (95% CI: 39.92-44.11, IQR: 39.92-44.11) for bilinear interpolation ($U = 1{,}847$, $p < 0.001$, Cohen's $d = 0.86$). Similarly, the SSIM indicator, proposed by J. Wang *et al.* (2004) to assess the structural similarity of images, showed a median value of 0.953 (95% CI: 0.935-0.968) versus 0.910 (95% CI: 0.893-0.927), respectively ($t = 12.47$, $df = 238$, $p < 0.001$, Cohen's $d = 1.71$). Correlation analysis revealed a strong positive correlation between PSNR and SSIM for both methods ($r = 0.847$ for convolutional neural networks, $r = 0.823$ for bilinear interpolation, $p < 0.001$), which confirms the internal consistency of quality metrics. As noted by J. Yang *et al.* (2023), for a more complete assessment of the quality of results, it is advisable to consider additional perceptual metrics that go beyond PSNR and SSIM. The distribution of results by PSNR category showed a clear advantage of convolutional neural networks in high-quality ranges (Table 1). Convolutional neural networks provided results with PSNR ⩾ 46 dB in 42.5% of cases, compared to 0% for bilinear interpolation. However, bilinear interpolation was more likely to show poor quality: 46.3% of the results had a PSNR < 42 dB versus 0.0% for convolutional neural networks. Analysis of the influence of content type revealed significant differences: for geometric shapes, the PSNR was 47.3 ± 2.1 dB (CNN) vs. 40.1 ± 2.5 dB (bilinear), for landscapes – 45.1 ± 2.4 dB vs. 41.8 ± 2.1 dB, for text images – 43.8 ± 2.8 dB vs. 43.1 ± 1.9 dB.

Table 1. Distribution of results over the PSNR range

PSNR range (dB)	Convolutional neural networks (%)	Bilinear interpolation (%)
<40	0.0	32.5
40-42	0.0	10.0
42-44	22.5	21.7
44-46	30.0	29.8
46-48	22.5	0.0
>48	19.2	0.0

Source: compiled by the authors

The most critical differences between the methods were observed in the accuracy of predicting geometric parameters. Bilinear interpolation showed perfect accuracy due to its deterministic nature: the median angle error was 0.0° (IQR: 0.0-0.0°), the maximum error was 0.0°, and the standard deviation was 0.0°. For convolutional neural networks, the median angle error was 47.95° (IQR: 28.33-61.24°, $M = 46.49°$, $SD = 23.72°$), which is statistically significantly worse ($U = 892.5$, $p < 0.001$). Regression analysis of the dependence of the error on the complexity of the transformation showed a statistically significant relationship for convolutional neural networks ($R^2 = 0.324$, $\beta = 0.569$, $p < 0.001$), whereas there was no such relationship for bilinear interpolation ($R^2 = 0.012$, $p > 0.05$). This indicates the sensitivity of neural networks to the complexity of input data and the need for additional training in more complex cases. A similar pattern was observed for the scaling error: the median scaling factor error for convolutional neural networks was 0.238 ($M = 0.261$, $SD = 0.144$) compared to the ideal accuracy (0.0) of bilinear interpolation ($U = 892.5$, $p < 0.001$). Thus, convolutional neural networks showed a systematic error in predicting both the rotation angle and the scaling factor, which confirms the fundamental nature of the discovered limitation of the neural network approach for problems of accurate prediction of geometric parameters.

Analysis of time characteristics revealed a complex performance pattern. The median processing time for convolutional neural networks was 447.0 ms with an interquartile span of 247.0-567.0 ms, while for bilinear interpolation, the median was 0.0 ms (many operations were performed instantly), but the average time was 358.9 ms with much greater variability ($SD = 421.1$ ms). The coefficient of variation in processing time for convolutional neural networks was 42.5% compared to 117.3% for bilinear interpolation, which indicates greater stability and predictability of the deep learning algorithm. Quantisation of the model to an 8-bit representation allowed achieving acceleration of infer without loss of accuracy. The performance of the Qualcomm Snapdragon 680 is in line with the benchmarks by V.J. Reddi *et al.* (2022), which confirms the representativeness of the results. A detailed analysis of the processing time distribution showed that 75% of convolutional neural network experiments were completed in less than 548 ms, whereas for bilinear interpolation, this figure was characterised by significant uncertainty due to the bimodal distribution. The maximum processing time for convolutional neural networks did not exceed 762.0 ms, while bilinear interpolation showed processing cases of more than 1,034 ms for complex transformations. A detailed comparison of key performance metrics is presented in Table 2.

Table 2. Comparative characteristics of geometric transformation methods

Metric	CNN	Bilinear interpolation	Relative difference (%)	Statistical significance
Median processing time (ms)	447.0 ± 177.0	358.9 ± 420.0	-19.8%	$U = 2,847$, $p < 0.0$
Median PSNR (dB)	45.56 ± 10.95	42.42 ± 2.17	+7.4%	$U = 1,847$, $p < 0.001$
Median SSIM	0.953 ± 0.225	0.910 ± 0.017	+4.7%	$t = 12.47$, $p < 0.001$
Median angle error (°)	47.95 ± 23.72	0.0 ± 0.0	---	$U = 892.5$, $p < 0.001$
Median scale error	0.238 ± 9.453	0.0 ± 0.0	---	$U = 892.5$, $p < 0.001$

Source: compiled by the authors

Monitoring of system resources showed that convolutional neural networks consumed an average of 127.3 ± 23.8 MB of RAM, compared to 89.2 ± 15.4 MB for bilinear interpolation ($t = 12.47$, $p < 0.001$). The CPU load was 67.8 ± 12.3% and 45.2 ± 18.7%, respectively. The battery's power consumption varied slightly: 18.3 ± 4.2 mAh versus 16.7 ± 3.9 mAh ($p > 0.05$). Analysis of the thermal mode of the device revealed an increase in the processor temperature by 3.2°C ± 1.8°C when using convolutional neural networks, compared to 1.9°C ± 1.1°C for bilinear interpolation. These differences can have practical implications for long image processing sessions. Comparison of the results obtained with the data of contemporary research confirms their representativeness and compliance with general trends in the industry. According to the benchmarking results by V.J. Reddi *et al.* (2022), the infer time for mobile neural networks on mid-segment chipsets is in the range of 300-600 ms for medium-complexity models, which is in good agreement with the obtained indicators (447.0 ms for CNN). The 127.3 MB memory consumption for the neural network also corresponds to the typical values for TensorFlow Lite-optimised models, as indicated in the review by T. Choudhary *et al.* (2020).

The discovered trade-off between visual quality and geometric accuracy was confirmed by the conclusions of Y. He & L. Xiao (2024) on the specifics of using neural network methods for various computer vision tasks. The authors of EfficientFormer also note that optimising the architecture for specific tasks is crucial for achieving the best results on mobile platforms. Contemporary optimisation approaches described by R. Archana & P.S.E. Jeevaraj (2024), can potentially improve the accuracy of geometric predictions through specialised model training. In particular, the knowledge distillation and neural architecture search methods open up prospects for creating models that combine high visual quality of the result with accurate prediction of geometric transformation parameters. The results obtained reveal a fundamental compromise between the quality of the

visual result and the accuracy of mathematical calculations of geometric parameters. Convolutional neural networks, while showing statistically significantly better PSNR and SSIM scores, also showed critically worse results in predicting accurate values of rotation angles and scaling factors. This phenomenon can be explained by various optimisation goals: convolutional neural networks are trained to maximise the visual similarity of the result to the reference, while bilinear interpolation provides mathematically accurate reproduction of the geometric transformation.

The results of the study confirmed the conclusions of A. Howard *et al.* (2019) on the possibility of efficient application of optimised convolutional architectures on mobile devices. Simultaneously, the identified limitations in the accuracy of predicting geometric parameters require additional attention when designing applications where mathematical accuracy is critical. The use of the compression and quantisation methods described by S. Han *et al.* (2015) and J. Yang *et al.* (2023), allows significantly improving performance on mobile devices without significantly losing the quality of visualisation. Practical recommendations are based on the identified patterns. For applications where the visual quality of the result is a priority (photo editors, social network filters), it is recommended to use convolutional neural networks based on MobileNet or ShuffleNet architectures. Bilinear interpolation remains the best choice for applications that require precise geometric calculations (navigation systems, medical diagnostics, industrial measurement). Hybrid approaches may be appropriate for augmented reality applications where both high visual performance and geometric accuracy are required.

Limitations of the study include the use of a single type of mobile device, which may limit the generalisability of the results to other hardware configurations. A fixed image size of 224×224 pixels may also not reflect the full range of practical applications. In addition, the convolutional neural network architecture used was not specifically optimised for geometric parameter prediction problems, which could affect the accuracy of the results. The results obtained are consistent with basic research in the field of image quality assessment and mobile neural networks. The image quality assessment was based on the classic work by J. Wang *et al.* (2004), where it was shown that human perception of quality correlates more with the structural characteristics of the image than with the pixel difference, which justifies the use of SSIM as the main metric in the conducted study.

J. Yang *et al.* (2023), in their review of methods for evaluating image quality based on deep learning, systematised contemporary approaches to automatic quality assessment. The researchers showed that convetional PSNR and SSIM metrics, although widely used, do not always correlate with human perception of quality for complex types of distortion. This highlights the need for a comprehensive assessment approach that was applied in this study. T. Choudhary *et al.* (2020) conducted a comprehensive review of model compression and acceleration methods. The researchers systematised approaches including pruning, quantification, knowledge distillation, and low-rank factorisation, showing the ability to achieve compression up to 50x while maintaining 95% accuracy. Further development in MobileNetV2 (Sandler *et al.*, 2018) with inverted residual blocks and MobileNetV3 (Howard *et al.*, 2019) with hardware-oriented architecture search continued the tradition of optimising mobile networks. The developed dual-input architecture is based on these principles of efficient design.

V.J. Reddi *et al.* (2022) presented MLPerf Mobile, a standardised benchmark for evaluating machine learning performance on mobile devices. The benchmark covers a variety of tasks, including image classification, object detection, and segmentation. The results of the study on the Motorola G32 with a Qualcomm Snapdragon 680 processor are consistent with the benchmark indicators for similar mid-range devices. M. Jaderberg *et al.* (2015) introduced Spatial Transformer Networks, which allow neural networks to explicitly model geometric transformations. This work demonstrated the possibility of learning end-to-end spatial transformations, which is conceptually close to the problem of predicting the parameters of geometric transformations in this study. However, the results obtained show that classical methods remain more reliable for accurate prediction of parameters.

Thus, the analysis of these studies indicates that contemporary approaches to the use of neural networks for image processing and geometric transformations demonstrate significant potential, especially in the context of mobile computing. Furthermore, the results of the experiment show that despite the conceptual proximity of deep models to the problem of estimating transformation parameters, in practical scenarios, classical algorithmic methods still provide higher accuracy and stability. This confirms the feasibility of further research aimed at combining the benefits of deep learning with proven classical approaches to improve the efficiency and reliability of computer vision systems on mobile devices.

CONCLUSIONS

As a result of the study, fundamental differences were established between convolutional neural networks and bilinear interpolation when performing geometric image transformations on mobile devices. Convolutional neural networks showed a statistically significant advantage in visual quality indicators: the median PSNR value was 7.4% higher (45.56 dB vs. 42.42 dB), and the SSIM was 4.5% higher (0.953 vs. 0.910) compared to bilinear interpolation. A critical problem of accuracy in predicting geometric parameters was identified: convolutional neural networks showed a median angle error of 47.95° compared to the ideal accuracy (0.0°) of bilinear interpolation.

Analysis of time characteristics revealed greater stability of convolutional neural networks: the coefficient of variation of processing time was 42.5% versus 117.3% for bilinear interpolation, which indicates better predictability of performance for interactive applications. The average processing time for convolutional neural networks was 447.0 ms, which is acceptable for most real-time mobile applications. It was found that the choice of the optimal method depends on the specific requirements of the application. The choice of the optimal method depends on the specific requirements of the application: for visual quality, it is advisable to use convolutional neural networks, for geometric accuracy – bilinear interpolation. Future research areas include the development of hybrid approaches that combine the advantages of both methods using optimisation techniques, testing across a wider range of mobile devices and processor architectures, and optimising convolutional neural network architectures specifically for the task of predicting geometric parameters whilst preserving visual quality.

ACKNOWLEDGEMENTS

The author expresses gratitude to the Department of Automation and Control in Technical Systems at the National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute" for providing technical resources and consultative support during the experimental research.

FUNDING

This research received no external funding. The study was carried out as part of a state-funded research project of the Department of Automation and Control in Technical Systems, Faculty of Electronics, at Igor Sikorsky Kyiv Polytechnic Institute, without additional funding.

CONFLICT OF INTEREST

None.

REFERENCES

[1] Archana, R., & Jeevaraj, P.S.E. (2024). Deep learning models for digital image processing: A review. *Artificial Intelligence Review*, 57(1), article number 11. doi: 10.1007/s10462-023-10631-z.

[2] Choudhary, T., Mishra, V., Goswami, A., & Sarangapani, J. (2020). A comprehensive survey on model compression and acceleration. *Artificial Intelligence Review*, 53(7), 5113-5155. doi: 10.1007/s10462-020-09816-7.

[3] Gou, J., Yu, B., Maybank, S.J., & Tao, D. (2021). Knowledge distillation: A survey. *International Journal of Computer Vision*, 129(6), 1789-1819. doi: 10.1007/s11263-021-01453-z.

[4] Han, S., Pool, J., Tran, J., & Dally, W.J. (2015). Learning both weights and connections for efficient neural network. In *Advances in neural information processing systems 28 (NeurIPS 2015)* (pp. 1135-1143). Montreal: Neural Information Processing Systems Foundation.

[5] He, Y., & Xiao, L. (2024). Structured pruning for deep convolutional neural networks: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(5), 2900-2919. doi: 10.1109/TPAMI.2023.3334614.

[6] Howard, A., *et al.* (2019). Searching for MobileNetV3. In *Proceedings of the IEEE/CVF international conference on computer vision (ICCV)* (pp. 1314-1324). Seoul: IEEE.

[7] Ignatov, A., *et al.* (2023). Efficient and accurate quantized image super-resolution on mobile NPUs, mobile AI & AIM 2022 challenge: Report. In L. Karlinsky, T. Michaeli & K. Nishino (Eds.), *Computer vision – ECCV 2022 workshops* (pp. 92-129). Cham: Springer. doi: 10.1007/978-3-031-25066-8_5.

[8] Jaderberg, M., Simonyan, K., Zisserman, A., & Kavukcuoglu, K. (2015). Spatial transformer networks. In *Advances in neural information processing systems 28 (NeurIPS 2015)* (pp. 2017-2025). Montreal: Neural Information Processing Systems Foundation.

[9] Li, Z., Li, H., & Meng, L. (2023). Model compression for deep neural networks: A survey. *Computers*, 12(3), article number 60. doi: 10.3390/computers12030060.

[10] Liu, H., Galindo, M., Xie, H., Wong, L.K., Shuai, H.H., Li, Y.H., & Cheng, W.H. (2024). Lightweight deep learning for resource-constrained environments: A survey. *ACM Computing Surveys*, 56(10), article number 267. doi: 10.1145/3657282.

[11] Mehta, S., & Rastegari, M. (2022). MobileViT: Light-weight, general-purpose, and mobile-friendly vision transformer. In *International conference on learning representations (ICLR 2022)* (pp. 1-26). Vancouver: ICLR.

[12] Menghani, G. (2023). Efficient deep learning: A survey on making deep learning models smaller, faster, and better. *ACM Computing Surveys*, 55(12). doi: 10.1145/3578938.

[13] Reddi, V.J., *et al.* (2022). MLPerf mobile inference benchmark: An industry-standard open-source machine learning benchmark for on-device AI. In *Proceedings of the Machine Learning and Systems Conference (MLSys 2022)* (pp. 352-364). Baltimore: PMLR.

[14] Rokh, B., Azarpeyvand, A., & Khanteymoori, A. (2023). A comprehensive survey on model quantization for deep neural networks in image classification. *ACM Transactions on Intelligent Systems and Technology*, 14(6), article number 97. doi: 10.1145/3623402.

[15] Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L.-C. (2018). MobileNetV2: Inverted residuals and linear bottlenecks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (CVPR)* (pp. 4510-4520). Salt Lake City: IEEE. doi: 10.1109/CVPR.2018.00474.

[16] Tan, M., & Le, Q.V. (2021). EfficientNetV2: Smaller models and faster training. In *Proceedings of the 38th international conference on machine learning* (pp. 10096-10106). Baltimore: PMLR.
[17] Wang, Z., Bovik, A.C., Sheikh, H.R., & Simoncelli, E.P. (2004). Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4), 600-612. doi: 10.1109/TIP.2003.819861.
[18] Wei, L., Ma, Z., Yang, C., & Yao, Q. (2024). Advances in the neural network quantization: A comprehensive review. *Applied Sciences*, 14(17), 744. doi: 10.3390/app14177445.
[19] Xiao, Z., Zhang, Z., Hung, K.-W., & Lui, S. (2021). Real-time video super-resolution using lightweight depthwise separable group convolutions with channel shuffling. *Journal of Visual Communication and Image Representation*, 75, article number 103038. doi: 10.1016/j.jvcir.2021.103038.
[20] Yang, J., Lyu, M., Qi, Z., & Shi, Y. (2023). Deep learning based image quality assessment: A survey. *Procedia Computer Science*, 221, 1000-1007. doi: 10.1016/j.procs.2023.08.080.

Порівняльний аналіз згорткових нейронних мереж та білінійної інтерполяції для мобільних застосунків

Алекс Мосиса

Аспірант
Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського»
03056, просп. Берестейський, 37, м. Київ, Україна
https://orcid.org/0009-0004-5845-6403

Ганна Власюк

Доктор технічних наук, професор
Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського»
03056, просп. Берестейський, 37, м. Київ, Україна
https://orcid.org/0000-0002-7382-0435

Анотація. Актуальність дослідження зумовлена зростаючою потребою в ефективних алгоритмах обробки зображень для мобільних пристроїв з обмеженими обчислювальними ресурсами. Мета роботи полягала у проведенні комплексного порівняння ефективності згорткових нейронних мереж та білінійної інтерполяції для передбачення параметрів геометричних трансформацій у мобільних застосунках. Дослідження проводилось на мобільному пристрої Motorola G32 з використанням dual-input архітектури згорткових нейронних мереж. Виконано 120 експериментів з зображеннями розміром 224×224 пікселів, які включали кути повороту від -45° до +45° та коефіцієнти масштабування від 0,5 до 1,5. Модель CNN оптимізована через TensorFlow Lite з 8-бітною квантизацією. Для оцінки якості зображень застосовано метрики PSNR та SSIM. Статистичний аналіз включав: критерій Шапіро-Вілка для перевірки нормальності, t-тест Стьюдента та U-критерій Манна-Уїтні для порівняння груп, коефіцієнт Коена для оцінки розміру ефекту, дисперсійний аналіз ANOVA, кореляційний аналіз Пірсона та Спірмена, регресійний аналіз. Здійснено моніторинг теплового режиму та споживання енергії пристрою при рівні значущості $p < 0,05$. Згорткові нейронні мережі демонстрували статистично значущо кращі показники якості зображення: медіанне значення PSNR становило 45,56 дБ порівняно з 42,42 дБ для білінійної інтерполяції, SSIM відповідно 0,953 та 0,910. Водночас білінійна інтерполяція забезпечувала ідеальну точність передбачення геометричних параметрів (медіанна похибка кута 0,0°) на відміну від згорткових нейронних мереж (медіанна похибка 47,95°). Час обробки для згорткових нейронних мереж становив 447,0 мс проти 358,9 мс для білінійної інтерполяції Споживання батареї: 18,3 ± 4,2 мАг (CNN) проти 16,7 ± 3,9 мАг (білінійна), $p > 0,05$. Температура процесора зростала на 3,2°C (CNN) проти 1,9°C (білінійна). Практична значимість. Розроблені методи можуть бути використані у мобільних застосунках доповненої реальності для оптимального вибору алгоритмів залежно від вимог застосунку

Ключові слова: глибоке навчання; обробка зображень; мобільні обчислення; параметри трансформації; метрики продуктивності; нейронна архітектура