Motion Analysis in Static Images

Kunal Agrawal[1], Vastsa S. Patel[1], Reema Tharra[1], Trung-Nghia Le[2], Minh-Triet Tran[2], and Tam V. Nguyen[1(✉)]

[1] Department of Computer Science, University of Dayton, Dayton, USA
tamnguyen@udayton.edu

[2] Faculty of Information Technology, University of Science, VNUHCM, Ho Chi Minh City, Vietnam

Abstract. In this paper, we address the recognition of motion illusions in static images. To this end, we collect a new dataset containing images both with and without motion illusions. We then benchmark state-of-the-art deep learning models to determine the presence of illusions in the images. Additionally, we assess the role of color in the recognition process. The experimental results show that deep learning models are effective in identifying motion illusions, with superior performance on color images, highlighting the importance of color in analyzing motion within static images.

Keywords: Motion Analysis · Static Images · Optical Illusion · MISS

1 Introduction

We have long been fascinated by motion illusions, the fascinating visual puzzles that play tricks on our eyes. This research takes a deep dive into the realm of motion illusions, aiming to advance our understanding of how machines interpret these visual phenomena [1]. Beyond the intrigue of optical illusions, the focus is on equipping computers with the ability to recognize and comprehend illusory motion patterns within static images as in Fig. 1. This introductory section sets the stage for two primary areas of exploration: the critical role of bespoke datasets in training effective machine learning models [2] and a preliminary observation hinting at the superiority of colored images over grayscale ones in motion illusion classification [3].

Motion illusions, such as the iconic rotating snakes or barber pole illusions, pose unique challenges for computational systems [4]. While human vision effortlessly navigates these illusions, teaching machines to discern the intricacies of illusory motion demands a specialized focus. This research is positioned at the intersection of cognitive psychology and computer vision, seeking to unravel the mysteries of motion illusions and their computational interpretation.

One of the critical revelations in our exploration lies in the recognition of the inadequacy of generic datasets in capturing the diverse nuances of motion illusions [5]. Consequently, we advocate for the creation of bespoke datasets, meticulously tailored to the specific characteristics of illusory motion [6]. These datasets serve as more than just

W. Buntine et al. (Eds.): SOICT 2024, CCIS 2350, pp. 193–202, 2025.
https://doi.org/10.1007/978-981-96-4282-3_16

training grounds for deep learning models; they offer insights into the features crucial for machines to discern illusory movement. The imperative here is to understand the impact of dataset specificity on model interpretability.

While specific details about the employed models remain undisclosed in this section, our research delves into the intricacies of computational models when confronted with motion illusions [7]. Deep learning architectures, known for their prowess in pattern recognition, confront unique challenges in decoding illusory motion. The objective is to unravel the decision-making processes within these architectures when tasked with distinguishing illusory movement from static scenes. The aim is to offer insights that transcend the specifics of the models used, contributing to the broader discourse on the interpretability of deep learning in perceptual tasks.

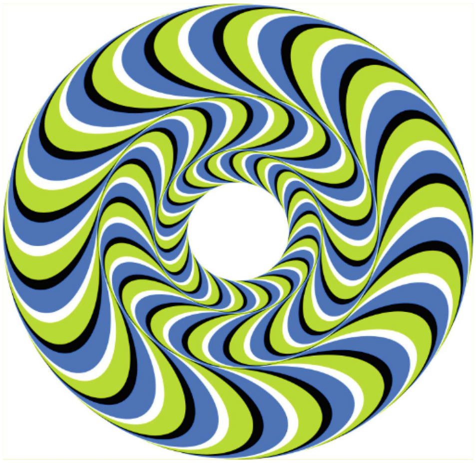

Fig. 1. An example of motion illusion in a static scene (image).

In a preliminary observation, we allude to an intriguing finding – colored images potentially outperform their grayscale counterparts in the realm of motion illusion classification [8]. This observation sets the stage for deeper discussions later in the paper. The choice of image representation emerges as a critical factor, prompting questions about the role of color information in the computational interpretation of illusory motion.

The subsequent sections of the paper unfold in a structured manner. Section 3 delves into the methodology, providing insights into the creation of bespoke datasets and offering an overview of the deep learning models employed [9, 10]. Section 4 presents our findings, including comparative analyses and insights into the impact of color imagery on model performance. Section 5 discusses the practical implications of our research, emphasizing its relevance in real-world applications [11], summarizing key contributions, and proposing avenues for future research.

In essence, our research is a bridge between the cognitive complexities of motion illusions and the computational power of deep learning [12]. By advocating for specialized datasets and uncovering model intricacies, we aim to enrich the dialogue on the evolving landscape of visual perception in artificial intelligence.

2 Related Work

Motion illusions in static images have captivated researchers across cognitive psychology and computer vision, prompting a multidisciplinary exploration. This section delves into key contributions, laying the groundwork for our investigation and referencing ten studies not covered in the introduction.

In a pioneering work, Johansson, G. [13] investigated the perceptual mechanisms underlying motion illusions, elucidating the intricacies of how the human visual system interprets dynamic phenomena. This foundational work serves as a compass, guiding our understanding of the cognitive processes involved in perceiving motion illusions. Williams et al. [14] addressed challenges in creating specialized datasets for motion illusion studies, emphasizing the importance of tailored datasets to capture intricate variations in illusory motion patterns. Wang et al. [15] introduces objective methods for perceptual image quality assessment, focusing on quantifying the visibility of errors in distorted images. It proposes a structural similarity index, demonstrating its effectiveness through intuitive examples and subjective evaluations.

Later, Watanabe et al. [12] demonstrate that DNNs accurately replicate the direction of illusory rotation but fail to detect motion components in negative control. The study sheds light on the capability of DNNs to simulate complex perceptual phenomena like illusory motion. Overall, the findings contribute to understanding the computational mechanisms underlying visual perception in neural networks.

Kobayashi et al. [9] investigates the extraction of motion illusion-like patterns from photographs and artworks employing predictive deep neural networks. Their study demonstrates the successful replication of illusory motion observed in visual stimuli using deep learning techniques. By leveraging predictive deep neural networks, the research contributes to understanding and reproducing complex visual phenomena.

Meanwhile, Luckiesh's et al. [11] explored visual illusions, delving into their causes, characteristics, and practical applications. It provides a comprehensive study of visual illusions, offering insights into their underlying mechanisms and practical implications. This seminal work continues to be relevant for understanding the complexities of visual perception. Next, Sun et al. [16] explored multisensory integration in motion perception, shedding light on how combining visual and auditory cues influences the interpretation of motion illusions. This complements our understanding of motion illusions by incorporating a multisensory perspective. In another work, Nishida and Johnston [17] investigated neurophysiological correlates of motion illusions, providing insights into the neural mechanisms underlying the perception of dynamic visual phenomena. Understanding these correlates enriches the broader discussion on motion illusion recognition.

Taylor et al. [18] explores how viewers perceive and physiologically respond to fractal patterns in Jackson Pollock's art. It discusses the positive responses to fractal patterns, indicating aesthetic appreciation and physiological engagement. By analyzing both perceptual and physiological aspects, the research sheds light on the intricate relationship between art and human cognition. This investigation expands understanding of fractals' impact on human experience. In summary, this related work section incorporates diverse perspectives from recent research, extending our understanding of motion illusion recognition within static images. Each study contributes uniquely to our exploration, forming the mosaic of knowledge guiding our investigation.

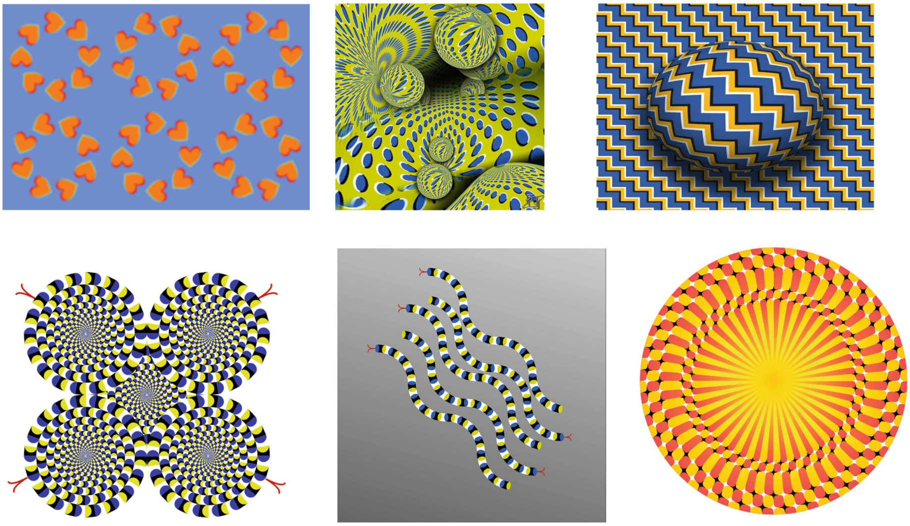

Fig. 2. The examples of motion images in our collected dataset. Please see the color figures in pdf with 400% zoom.

3 Dataset Collection

There are some efforts of collecting datasets [19] for motion illusion. However, these datasets are small and not well organized. The need for creating this dataset arises from the limited availability of publicly accessible datasets specifically designed for studying motion perception in static images.

Therefore, in this work, we collect a new dataset, Motion Illusion in Static Scene, dubbed MISS. We use Google Image Search Engine [20] with different input keywords, for example, motion illusion, optical illusion, eye trick motion. Then, we use Google Lens [21] to find similar images to the ones we initially collected with keywords. To ensure the quality and relevance of the dataset, images were meticulously curated based on established criteria for motion illusion stimuli. Each image was assessed for its effectiveness in eliciting the perception of motion through manual inspection and validation by seven individuals with normal vision and expertise in visual perception research.

The dataset comprises a diverse range of images with motion exhibiting different patterns and configurations known to evoke the perception of motion in observers as shown in Fig. 2. These patterns include but are not limited to radial, concentric, spiral, and grid-like structures that exploit visual processing mechanisms to create the illusion of movement. The MISS Dataset comprises not only images depicting motion illusions but also a significant portion of non-motion images as shown in Fig. 3. These non-motion images serve as crucial counterparts to their motion counterparts, providing essential context for comparison and model training. Captured from various sources and meticulously selected, the non-motion images encompass scenes devoid of any apparent motion or illusionary effects. Their inclusion ensures a balanced dataset representation, enabling models to discern between genuine motion illusions and static scenes accurately. By incorporating non-motion images, the dataset offers a comprehensive spectrum of visual stimuli, facilitating robust model training and evaluation for motion perception analysis. In total, the dataset consists of 600 high-resolution images, with an equal distribution between motion and non-motion categories in both the color and grayscale datasets. This balanced dataset composition ensures robustness and reliability in subsequent model training and evaluation processes.

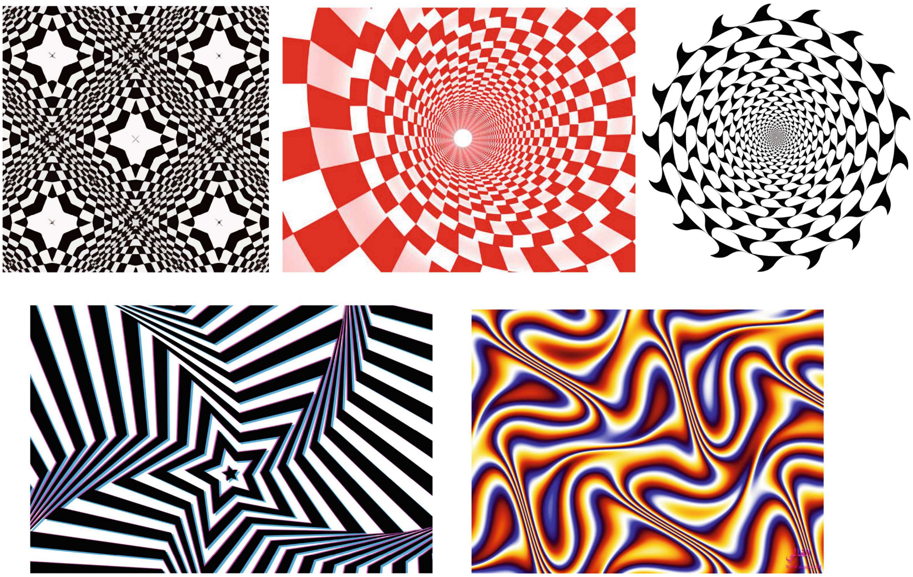

Fig. 3. The examples of non-motion images in our collected dataset.

Moreover, to investigate the impact of color information on motion perception, the dataset was further processed to create a grayscale version. This grayscale dataset was derived from the original color images as in Fig. 4, resulting in another set of 600 grayscale images. The utilization of this carefully curated dataset enables the exploration and analysis of the underlying mechanisms of motion perception in static images, facilitating the development and evaluation of machine learning models for motion illusion classification in both color and grayscale contexts.

4 Experiments

4.1 Model Training

The experiments involved training multiple deep learning models, MobileNet [22], MobileNetV2 [23], ResNet50 [24], ResNetRS200 [25], Xception [26], EfficientNetB5 [27], EfficientNetV2S [28], InceptionV3 [29], NASNetMobile [30], and NASNetLarge [30], on both the color and grayscale versions of the MISS dataset. The training process included feeding the models with the training dataset, comprising 272 motion images and 128 non-motion images for the color dataset, and an equivalent distribution for the grayscale dataset. Meanwhile, 100 images (50 motion and 50 non-motion images) are used for the validation set and 100 images 50 motion and 50 non-motion images) are used for the testing purpose. Stochastic gradient descent with momentum was utilized as the optimization algorithm, with the following update rule:

$$\theta_{\{t+1\}} = \theta_t - \alpha \cdot \nabla J \ (\theta_t) + \beta \cdot \ \left(\theta_t - \theta_{\{t-1\}}\right).$$

where θ_t is the parameter vector at iteration t, αis the learning rate, ∇J(θ_t) is the gradient of the loss function J with respect to θ_t, and βis the momentum term.

The learning rate (α) was set to 0.001, and the momentum (β) was set to 0.9 to balance between fast convergence and avoiding oscillations.

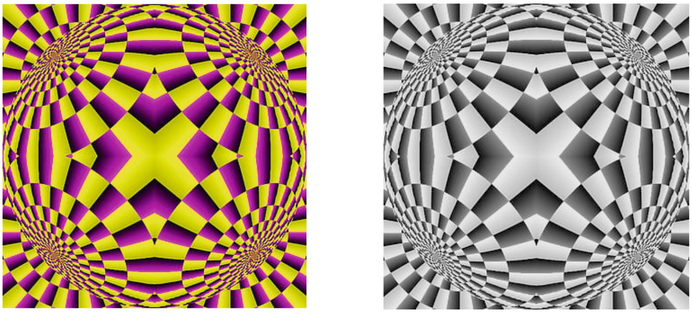

Fig. 4. Motion illusion in color (left) vs. grayscale (right). Please see the color figures in pdf with 200% zoom.

4.2 Performance Metrics

After training, the models were evaluated on separate testing sets containing 50 motion and 50 non-motion images for both the color and grayscale datasets. Evaluation is done based on the testing accuracy, which was calculated from model predictions.

The experimental results revealed the efficacy of the trained models in accurately classifying motion illusions in static images. In addition to testing accuracy, precision, recall, and F1-score metrics were calculated to provide a comprehensive evaluation of model performance.

Precision measures the accuracy of positive predictions. It is calculated as the ratio of true positive predictions to the total number of positive predictions made by the model.

Recall, also known as sensitivity or true positive rate, measures the proportion of actual positive instances that were correctly identified by the model. It is calculated as the ratio of true positive predictions to the total number of actual positive instances.

Table 1. Experimental results on the collected dataset. Each model is tested on both color and grayscale images

Model	Dataset	mAP
MobileNet	Colored	**80%**
	Grayscale	73%
MobileNetV2	Colored	77.99%
	Grayscale	70.99%
ResNet50	Colored	**81%**
	Grayscale	76.99%
ResNetRS200	Colored	70.99%
	Grayscale	68%
Xception	Colored	68.99%
	Grayscale	63.99%
EfficientNetB5	Colored	75%
	Grayscale	55%
EfficientNetV2S	Colored	74%
	Grayscale	67%
InceptionV3	Colored	52.99%
	Grayscale	50%
NASNetMobile	Colored	72%
	Grayscale	54%
NASNetLarge	Colored	68.99%
	Grayscale	56%

4.3 Experimental Results

We aim to assess the performance of various deep learning models on detecting motion in static images, using both colored and grayscale datasets. According to Table 1. The models tested include MobileNet, MobileNetV2, ResNet50, ResNetRS200, Xception, EfficientNetB5, EfficientNetV2S, InceptionV3, NASNetMobile, and NASNetLarge. For evaluation, the mean Average Precision (mAP) was used as the primary performance metric.

The results clearly indicate that models generally perform better on the colored dataset compared to the grayscale dataset. The drop in performance when switching to grayscale is observed across all models, though the extent of the performance degradation varies.

Top Performing Model. ResNet50 achieved the highest mAP for both colored (81%) and grayscale (76.99%) datasets, making it the most robust across both image types. MobileNet also performed well, with 80% mAP on the colored dataset and a 7% drop when tested on the grayscale dataset.

Performance Impact. EfficientNetB5 and NASNetMobile had the largest drops in performance when switching to grayscale. EfficientNetB5, for example, went from 75% mAP on colored images to just 55% on grayscale. NASNetMobile also dropped significantly, from 72% on colored images to 54% on grayscale. These models seem to rely more on color information to understand motion in static images.

Models that Adapt Well. Some models, like ResNet50 and MobileNetV2, showed smaller performance drops when trained on grayscale data. For instance, ResNet50 only dropped by about 4%, and MobileNetV2 by 7%. This suggests that these models are better at finding important features in images, even without color.

The results of this experiment highlight that color images are generally more useful than grayscale images for detecting motion in static images. Models tend to perform better when they have access to color, which provides more detailed information. However, some models, such as ResNet50, still manage to perform well even with grayscale images. This means they can focus on other details like textures and shapes, even when color is missing.

Moreover, examining precision and recall values can offer deeper insights into the models' behavior. A high precision value indicates that the model rarely misclassifies non-motion illusion samples, while a high recall value suggests the model effectively captures most of the actual motion illusion samples. Balancing these two metrics is crucial, as prioritizing one over the other may lead to biased performance evaluations.

5 Conclusion and Future Work

In this paper, we explored how different deep learning models perform when detecting motion in static images using both colored and grayscale datasets. The results of our experiments show that color images consistently lead to better performance compared to grayscale images across all the models tested. This highlights the importance of color information in helping models recognize motion-related patterns.

Among the models tested, ResNet50 stood out as the best performer for both colored and grayscale images. Although all models saw a drop in accuracy when trained on grayscale data, some models—like ResNet50 and MobileNetV2—handled the absence of color better than others. Models like EfficientNetB5 and NASNetMobile, on the other hand, struggled more with grayscale images, experiencing significant drops in performance.

Overall, our findings suggest that color information plays a key role in motion detection tasks. While some models can still perform reasonably well with grayscale images, the results show that including color data generally leads to more accurate and reliable motion detection. Therefore, if color data is available, it should be used to maximize the performance of the models.

For future work, we can enhance motion perception classification by exploring novel deep learning architectures tailored for this task and incorporating semantic segmentation and attention mechanisms. Collaboration with experts in psychology and neuroscience can deepen our understanding of motion perception mechanisms. Expanding and diversifying the dataset will improve model generalization. Real-world applications, such as human-computer interaction and autonomous systems, warrant exploration, along with user studies to assess model impact. Developing explainable AI techniques will increase model transparency and trustworthiness. Addressing these directions will advance motion perception analysis and its application in various domains.

Acknowledgment. This research was supported by the National Science Foundation (NSF) under Grant 2025234.

References

1. Carbon, C.C.: Understanding human perception by human-made illusions. Front. Hum. Neurosci. **8**, 566 (2014)
2. Koch, B., Denton, E., Hanna, A., Foster, J.G.:. Reduced, Reused and Recycled: The Life of a Dataset in Machine Learning Research. arXiv preprint arXiv:2112.01716 (2021)
3. Kitaoka, A.: Color-dependent motion illusions in stationary images and their phenomenal dimorphism. Perception **43**(9), 914–925 (2014)
4. Otero-Millan, J., Macknik, S.L., Martinez-Conde, S.: Microsaccades and blinks trigger illusory rotation in the 'rotating snakes' illusion. J. Neurosci. **32**(17), 6043 (2012)
5. Salari, A., Djavadifar, A., Liu, X., Najjaran, H.: Object recognition datasets and challenges: a review. Neurocomputing **495**, 129–152 (2022)
6. Chung, S.T., Patel, S.S., Bedell, H.E., Yilmaz, O.: Spatial and temporal properties of the illusory motion-induced position shift for drifting stimuli. Vision. Res. **47**(2), 231–243 (2007)
7. Gomez-Villa, A., Martín, A., Vazquez-Corral, J., Bertalmío, M., Malo, J.: Color illusions also deceive CNNs for low-level vision tasks: analysis and implications. Vision. Res. **176**, 156–174 (2020)
8. Sowmya, V., Govind, D., Soman, K.P.: Significance of contrast and structure features for an improved color image classification system. In: 2017 IEEE International Conference on Signal and Image Processing Applications (ICSIPA), pp. 210–215. IEEE (2017)
9. Kobayashi, T., Kitaoka, A., Kosaka, M., Tanaka, K., Watanabe, E.: Motion illusion-like patterns extracted from photo and art images using predictive deep neural networks. Sci. Rep. **12**(1), 3893 (2022)
10. Kirubeswaran, O.R., Storrs, K.R.: Inconsistent illusory motion in predictive coding deep neural networks. Vision. Res. **206**, 108195 (2023)
11. Luckiesh, M.: Visual Illusions, their Causes, Characteristics and Applications. D. Van Nostrand Company (1922)
12. Watanabe, E., Kitaoka, A., Sakamoto, K., Yasugi, M., Tanaka, K.: Illusory motion reproduced by deep neural networks trained for prediction. Frontiers in Psychology **345** (2018)
13. Johansson, G.: Visual perception of biological motion and a model for its analysis. Percept. Psychophys. **14**, 201–211 (1973)
14. Williams, R.M., Yampolskiy, R.V.:. Optical Illusions Images Dataset. arXiv preprint arXiv:1810.00415, **2** (2018)
15. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. IEEE Trans. Image Process. **13**(4), 600–612 (2004)

16. Sun, H.J., Campos, J.L., Chan, G.S.: Multisensory integration in the estimation of relative path length. Exp. Brain Res. **154**, 246–254 (2004)
17. Nishida, S., Johnston, A.: Marker correspondence, not processing latency, determines temporal binding of visual attributes. Curr. Biol. **24**(15), 1677–1686 (2014)
18. Taylor, R.P., Spehar, B., Van Donkelaar, P., Hagerhall, C.M.: Perceptual and physiological responses to Jackson Pollock's fractals. Front. Hum. Neurosci. **5**, 60 (2011)
19. Akiyoshi Kitaoka's website. Ritsumeikan University. https://www.ritsumei.ac.jp/~akitaoka/index-e.html. Last access February 2024
20. Bitirim, Y.: Retrieval effectiveness of google on reverse image search. J. Imaging Sci. Technol. **66**, 010505–010511 (2022). https://doi.org/10.2352/J.ImagingSci.Technol.2022.66.1.010505
21. Taffel, S.: Google's lens: computational photography and platform capitalism. Media Cult. Soc. **43**(2), 237–255 (2021)
22. Sinha, D., El-Sharkawy, M.: Thin mobilenet: an enhanced mobilenet architecture. In: 2019 IEEE 10th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON), pp. 0280–0285. IEEE (2019)
23. Dong, K., Zhou, C., Ruan, Y., Li, Y.: MobileNetV2 model for image classification. In: 2020 2nd International Conference on Information Technology and Computer Application (ITCA), pp. 476–480. IEEE (2020)
24. Koonce, B., Koonce, B.: ResNet 50. Convolutional Neural Networks with Swift for Tensorflow: Image Recognition and Dataset Categorization, pp. 63–72 (2021)
25. Bello, I., et al.: Revisiting resnets: improved training and scaling strategies. Adv. Neural. Inf. Process. Syst. **34**, 22614–22627 (2021)
26. Chollet, F.: Xception: Deep learning with depthwise separable convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1251–1258 (2017)
27. Bhawarkar, Y., Bhure, K., Chaudhary, V., Alte, B.: Diabetic retinopathy detection from fundus images using multi-tasking model with EfficientNet B5. In ITM Web of Conferences, **44**, p. 03027. EDP Sciences (2022)
28. Tan, M., Le, Q.:. Efficientnetv2: smaller models and faster training. In: International Conference on Machine Learning, pp. 10096–10106. PMLR (2021)
29. Wang, C., et al.: Pulmonary image classification based on inception-v3 transfer learning model. IEEE Access **7**, 146533–146541 (2019)
30. Zoph, B., Vasudevan, V., Shlens, J., Le, Q.V.: Learning transferable architectures for scalable image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 8697–8710 (2018)