

Research Article

Deep Learning-Based Business Recommendation System in Intelligent Vehicles

Kyungho Yu,¹ Kiho Lim ,² and Pankoo Kim ¹

¹Department of Computer Engineering, Chosun University, Gwangju 61452, Republic of Korea

²Department of Computer Science, William Paterson University of New Jersey, Wayne, NJ 07470, USA

Correspondence should be addressed to Kiho Lim; limk2@wpunj.edu

Received 17 January 2022; Revised 7 September 2022; Accepted 7 October 2022; Published 14 April 2023

Academic Editor: Floriano Scioscia

Copyright © 2023 Kyungho Yu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The advancements in intelligent vehicle technologies are facilitating the growth of information technology (IT) platforms, unlike conventional automobiles. In-vehicle-infotainment (IVI) is becoming an appealing element in intelligent vehicles as it offers various experiences to users; however, it requires personalized services to provide even more sophisticated user experiences. It is supposed that passengers search for businesses that provide products or services they found interesting in videos played via IVIs while the vehicle is driving autonomously. In that case, it could be more effective to use images that can express the user's preference as a query for the search than to utilize texts such as product names. Accordingly, this study proposes a recommendation system that informs users of businesses near an intelligent vehicle when a passenger inputs an image of a product or service into an IVI system. The proposed recommendation system involves training deep learning-based image classification models with the user's interest images to classify the category, measure the similarity with the business category using Word2vec, and finally provide the locations of the businesses with a high degree of similarity via IVI, using a smartphone. The experimental results indicated that the user's interest image exhibited 85% accuracy for category classification via the EfficientNet B0 model, while the similarity between the image and business categories using Word2vec was particularly high in the business category similar to the actual image category.

1. Introduction

Having the advancements in 5G network technology and artificial intelligence, significant progress has also been made in the Internet-of-things (IoT) technology connecting objects with other objects, including intelligent vehicles, which can provide various services to drivers by connecting them to the Internet. Intelligent vehicles are always connected to the Internet, which facilitates vehicle-to-vehicle and vehicle-to-infrastructure communications. Intelligent vehicles are growing as one of the information technology (IT) platforms on which various content and services can be enjoyed via in-vehicle-infotainment (IVI) systems rather than simple transportation means such as conventional automobiles [1–3].

For intelligent vehicles to appeal to users based on necessity, services conveying convenience and comfort can be provided via IVI systems [2, 3]. Moreover, IVI enables

driving and traveling to be more enjoyable in intelligent vehicles with entertainment activities such as listening to music or watching movies. Consequently, IVI has become a critical marketing element for automakers when consumers purchase intelligent vehicles [2]. To offer a variety of experiences to consumers, automakers provide developers with software development kits (SDKs) appropriate for the operating systems (OS) of their vehicles which infotainment-using smartphones are being developed by software companies such as Apple and Google. Studies on IVI services have been conducted on accident prevention based on the driver's heart rate [4], video streaming [5], music recommendation [6], and vehicle maintenance [7]. In addition, automobile and software companies, including academia, have been attempting to provide a wide range of experiences and comfort to users via the IVI systems in intelligent vehicles. In the future, IVIs must focus on

personalization and context awareness to enhance user experience, going beyond providing various experiences to users in intelligent vehicles [3].

In intelligent vehicles capable of level 3 or higher automation driving, passengers can perform secondary tasks using IVIs, rather than driving [8]. Such vehicles can autonomously drive on expressways or parks, thus allowing their drivers or passengers to watch videos, read books, or go on social media with IVIs. While watching videos in an intelligent vehicle that can drive autonomously, passengers may decide to purchase furniture or an item they see in the video, but they may not know the exact name to search as a keyword when searching for nearby stores. Furthermore, while surfing the Internet or going through social media, passengers may find certain meals or hairstyles they like, but they may not know the exact search word for an appropriate restaurant or hair salon. Accordingly, infotainment services are required to provide assistance when passengers are unaware of the exact keyword to use when searching for nearby stores for a service or an item they discover while using IVIs.

The easiest way to find restaurants or stores to purchase an item in an intelligent vehicle is to adopt IVI's navigation system. The navigation system in intelligent vehicles converts the voice into text based on automatic speech recognition, and then the converted text undergoes natural language processing to recommend destination points by identifying the destination and point-of-interest from the user's input text [9]. However, the navigation system of IVIs using text cannot provide sufficient information on a product or a service to a user; hence, it is difficult to navigate to the user's desired destination. As aforementioned, conventional IVI navigation systems cannot be used if passengers do not know the exact name of a product or a service. If the text cannot be used as a query when using the IVI's navigation system, an image-based location search can recommend better destinations than a text-based search because an image can better represent user preferences, which cannot be reflected in texts when an image is used as a query [10]. Therefore, a system is required to recommend nearby businesses by using the image of a product or service as a query when the passenger of an intelligent vehicle decides to utilize the navigation system. Existing deep learning-based recommendation systems have exhibited high performance across several recommendation systems, including product, place, and movie recommendations. However, these systems have certain limitations, requiring multiview information such as user preferences (ratings and clicks) and various attributes (image, description, and reviews). The multiview information integration is a compute-intensive task. In addition, the main challenge in such integration is to retain information relevant to prediction and minimize other irrelevant information, which can reduce a large amount of data, time, and cost requirements when making recommendations. Given the wide availability of image data and the state-of-the-art performance of DL models for classification, we proposed a business location recommendation system that takes only image data of a product or service category desired by a passenger in an intelligent vehicle. The proposed method uses a DL model to

identify a product or service category and recommends nearby businesses that fit better into the classified category. Therefore, using image data only to express a user's preference as input and posing the recommendation problem as an image classification task mitigates the limitations mentioned above of existing recommendation systems.

The goal of this study is to propose a recommendation system that provides the user with a peripheral business location suitable for the product category based on the image entered into the IVI system. The process of the proposed IVI business location recommendation system is shown in Figure 1. The proposed recommendation system consists of two components: the user's interest in image classification and the similarity measurement between an input image and business categories. In an Intelligent vehicle, the passenger enters a picture of a store or service into the IVI system that they want to find. The input image goes through a deep learning-based image classification model to predict the input image category. Then, the similarity is measured between the word embedding of the category and the business category of the image. Once the similarity is measured, nearby stores with high similarity are navigated via IVI, using the smartphone.

The remainder of this paper is organized as follows. Related studies on deep-learning based recommendation systems and infotainment for intelligent vehicles are reviewed in Section 2. Then, the proposed image-based business recommendation system is presented in Section 3. Subsequently, the results obtained from experimenting with the collected images and relevant analysis are provided in Section 4. Finally, the conclusion of this study and possible future research directions are provided in Section 5.

2. Related Works

2.1. Infotainment for Intelligent Vehicles. Infotainment is a compound word for information and entertainment. IVI is a significant factor for consumers to purchase intelligent vehicles [2]. Existing car infotainment is divided into methods that either use the service built into the car embedded OS or connect a smartphone to the vehicle. In order to provide more services to users through IVI, an automaker that makes intelligent vehicles is putting a lot of effort into creating applications by providing SDK (soft development kit) suitable for the company's OS to developers. Infotainment using smartphones is led by software companies such as Apple-Apple CarPlay [11] and Google-Android Auto [12].

Existing research on IVI can be grouped into research on convenient operation and infotainment applications in using infotainment in intelligent vehicles. As an example of a convenient process, the authors in [13] have proposed a text-based infotainment system that takes voice command as input and uses Kaldi and voice recognition tools to change the command into text for running a service. In [14, 15], the authors have proposed an alternative infotainment system in which a driver can control the system by using gestures. The system is powered by a computer vision-based gesture recognition system. In research on infotainment applications, there is a study on using a wearable device to check

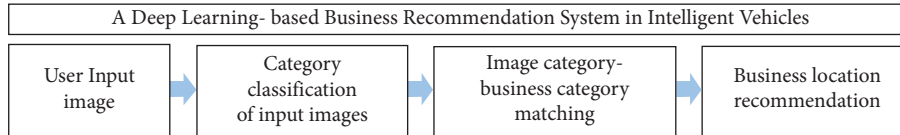


FIGURE 1: Process of the proposed business-location recommendation system.

heart rate and monitoring through infotainment to prevent a vehicle crash due to a stroke or heart attack while driving [4], and smooth multimedia streaming in an intelligent vehicle in operation. A study [5] protects the appropriate data rate and frames for this purpose. And, there is a study that recommends music by inputting the driver’s profile, current situation, and personal preference as inputs [6]. The vehicle maintenance system “e-talk” can check the vehicle’s condition through infotainment using the IoT sensors attached to the vehicle [7].

2.2. Recommendation System Based on Deep Learning. Recently, research is actively being conducted on applying deep learning (DL) techniques to conventional recommendation systems such as collaborative or content-based filtering [16]. For DL-based collaborative filtering, studies have been conducted on giving recommendations to a user by combining two different latent vectors after inputting information (including the descriptions, reviews, and ratings of a product or a place for embedding) of another user in the network who has similar preferences as that user [17, 18]. Information on a product or place is expressed as latent factors of the same dimension through a trained embedding layer and training in a neural network. The main advantage of such representation is that more complicated interactions can be captured in the categorical data. Similarly, in another study, product information such as images, descriptions, and review text were input into an auto-encoder to extract its features, and then the extracted features were integrated with the last layer to make recommendations to users [19]. The advantages of a DL-based recommendation system are that features can be extracted using the user’s preferences (review, rating, and clicks) and various attributes (image, description, and review) from a large amount of data. In addition, detailed recommendations can be provided via the complicated interactions between the extracted features [20].

Existing deep learning-based recommendation systems have exhibited high performance across several recommendation systems, including product, place, and movie recommendations; however, these systems are posed with certain limitations as they require multiview information. The main challenge is to retain information relevant to prediction and minimize other irrelevant information. The existing systems require a large amount of data, time, and cost to make detailed recommendations [21]. Therefore, unlike existing recommendation systems that employ large-scale data and complex deep learning models, the proposed business location recommendation system classifies the image category of a product or service desired by a passenger in an intelligent vehicle using a deep learning model and recommends nearby businesses that fit into the classified category.

3. Methods

3.1. Businesses Recommendation Systems in Intelligent Vehicles. This section describes the overall process of the system for recommending businesses near a vehicle when a passenger inputs an image of interest for a product or service into IVI within intelligent vehicles, as illustrated in Figure 2. First, a user’s interest category is selected to speculate the product or service image the user may be interested in. The range of individuals’ interest products or services is substantially wide; hence, local business data provided by Yelp were adopted to limit the range of the user’s interest categories. Each store in Yelp’s local business data comprises multiple categories, and among them, a category is selected as the user’s interest category, provided the frequency of appearance is high. The dataset is constructed by collecting images with the selected category as the web search keyword. Subsequently, the category classification of interest images is performed by training a deep learning-based image classification model with the user’s interest image dataset. To recommend appropriate businesses to users, the image category and business data need to be matched. The similarity between image and business categories is mapped in the same dimension using Word2vec and then measured accordingly. Once the similarity is measured between the two categories, the location of a business near the intelligent vehicle boarded by the passenger is provided to users via IVI, using a smartphone.

In Sections 3.2 and 3.3, image classification via deep learning and text similarity measurement using Word2vec, which are crucial parts of the proposed recommendation system, is explained.

3.2. Image Classification Model for Business Category. Deep learning models were adopted in this study to extract visual information from the image input by a user into the infotainment system of intelligent vehicles and to classify business categories. Using the ImageNet dataset, deep learning models for image classification were trained by applying transfer learning of the collected user’s interest category data to pretrained ResNet [21], Inception v3 [22], and EfficientNet B0 [23]. The architecture of each model is illustrated in Figure 3. Figure 3(a) presents the architecture of ResNet. Generally, it is expected that the performance of a deep learning model improves as the number of hidden layers in the model increases. However, deeper DL models suffer from vanishing gradient problems, and the use of pooling layers puts restrictions on the depth of the model. ResNet adopts a skip connection where the layer input is directly connected to the layer output to address such problems (Figure 3(a)). Using skip connection address the

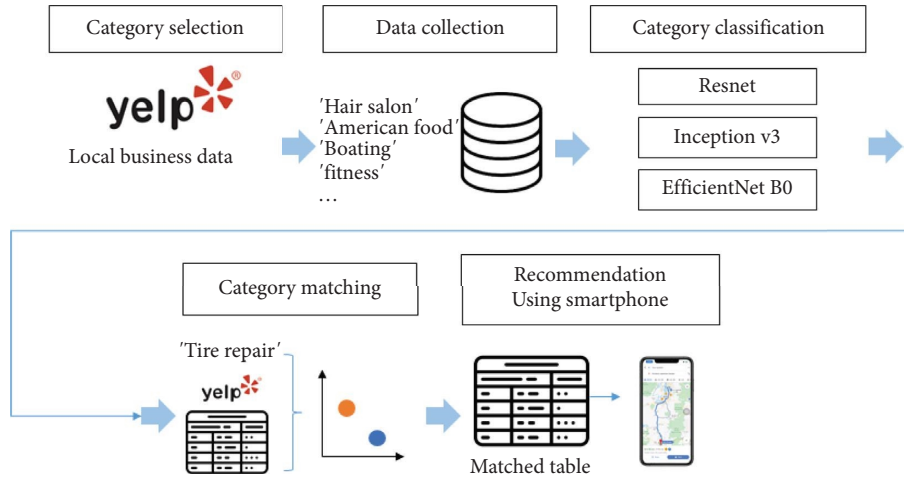


FIGURE 2: Business location recommendation process.

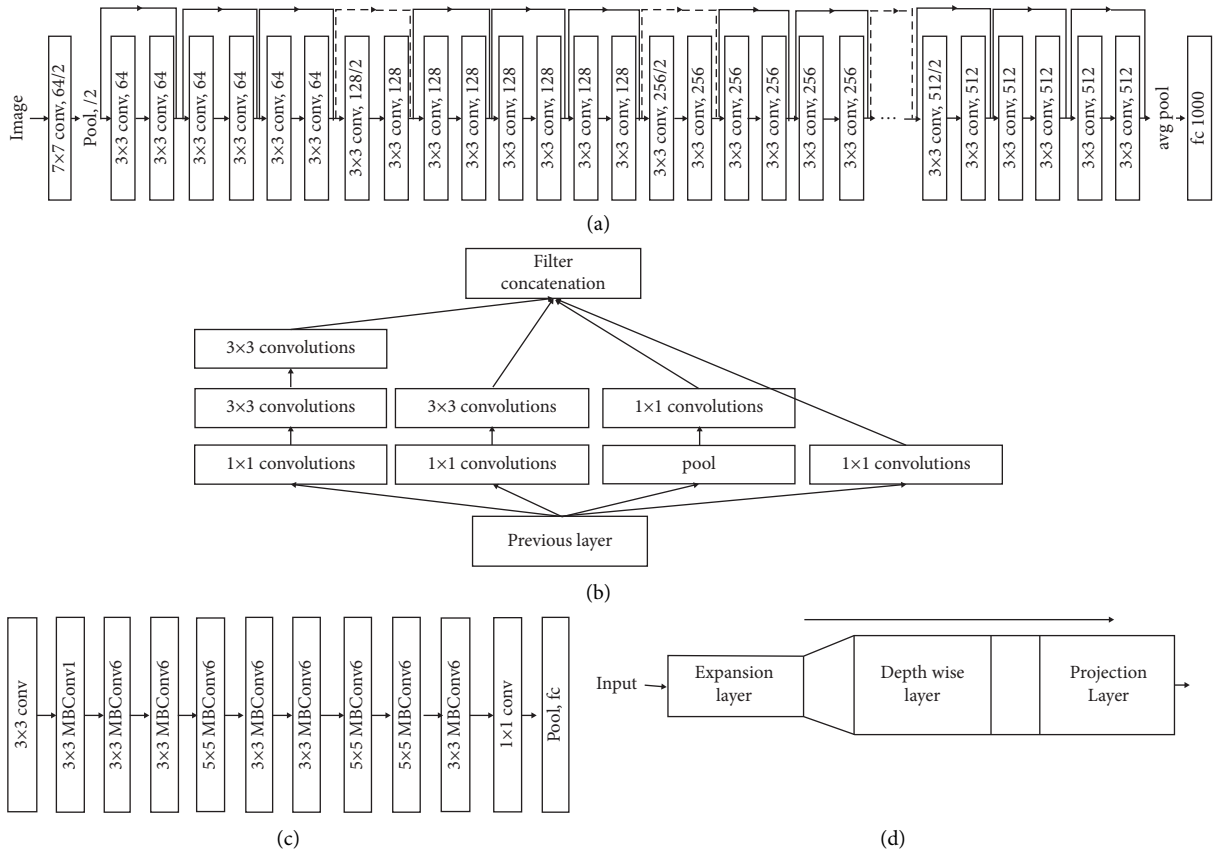


FIGURE 3: Architecture of (a) ResNet, (b) Inception v3 module, (c) EfficientNet B0, and (d) MBConv.

performance degradation challenges even when layers get deeper as the neural network learns the difference between input and output values. Figure 3(b) illustrates the module of the Inception v3 model. Inception v3 is a convolution neural network comprising 48 layers, where the model is configured by stacking multiple inception modules with the inception model presented in Figure 3(b) as the base. One module is configured using convolution filters of different sizes, unlike AlexNet or VGGNet models, which utilize the filter of the

same size in one layer for convolution. More diverse features can be extracted from the input image because the filters with different convolution sizes are adopted, and the computation amount can be reduced in the inception model by performing 1×1 convolution to reduce the number of parameters to be delivered to the convolution filter of a large size. For EfficientNet, a scale-up method required for improving the performance of the existing ConvNet, was investigated. Experiments were conducted on three scale-up

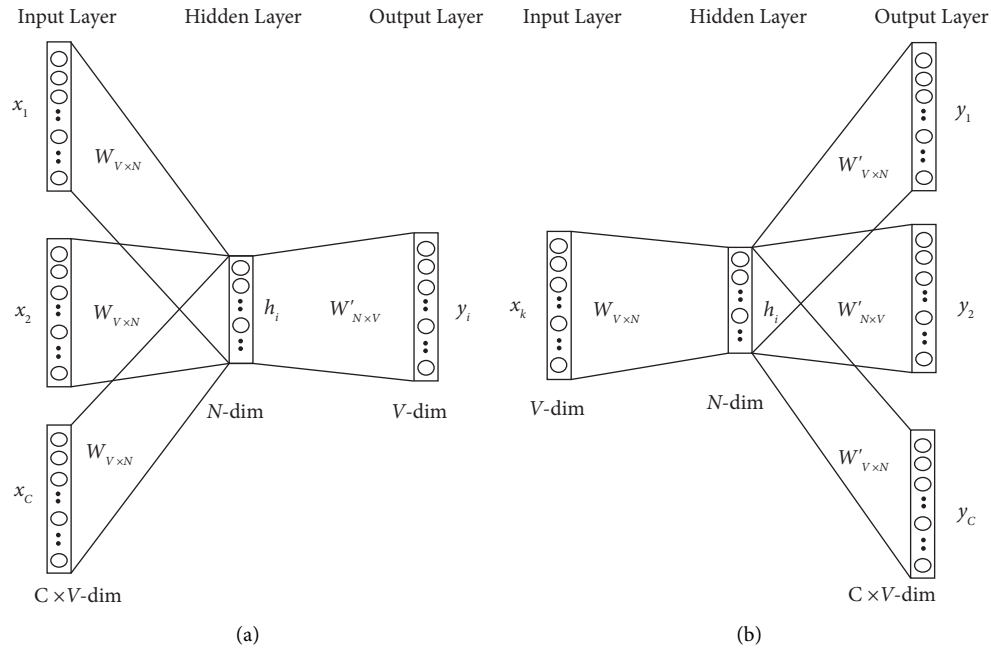


FIGURE 4: Architecture of Word2vec. (a) CBOW. (b) Skip-gram.

methods, which include increasing the depth of a neural network, increasing the channel width, and increasing the resolution of an image for training. In addition, the EfficientNet B0 model was created using the MBConv (Figure 3(d)) of MobileNet. MBConv is a mobile inverted bottleneck convolution adopted in MobileNet. This block expands the low-dimensional input to a 1×1 layer, expands it again into 3×3 convolution, and the outputs using the 1×1 convolution layer. This configuration is adopted because more features can be extracted by converting from the low- to high-dimensional data and using the projection layer while reducing the computation amount. EfficientNet exhibited high performance in image classification by adopting scale-up methods and MBConv. In this study, the user's interest data collected via web crawling were used to train ResNet, Inception v3, and EfficientNet B0 models to identify an appropriate model for the collected data.

3.3. Similarity Measurement Using Word2vec. Word embedding refers to converting text into numbers via distributed representation to enable a computer to understand human language. In other words, words comprising the text are mapped to real number vectors. Word2vec can convert text into vectors by identifying the relationship between words in a sentence and can perform operations such as addition or subtraction between words using the converted vectors. Figure 4 presents the architecture of Word2vec, which inputs the text into a thin neural network for training. The training methods of Word2vec include a continuous bag of words (CBOW), which predicts the center word based on the surrounding words, and the Skip-gram, which predicts surrounding words based on the center word. When compared, the Skip-gram model exhibited higher semantic

and syntactic accuracies than the CBOW model. The similarity was measured using the word embedded vectors of image and business categories because categories other than the business categories are ignored if the image category simply determines whether it is included in the business category when matching the categories. For example, if the image input by the user is classified as "American food," the categories of restaurant A are "wine bars, restaurants, food, beer, seafood, and steak houses." The similarity between these two has no correlation if the text is simply matched with the words "American food;" however, the degree of similarity can be identified if Word2vec, which learns the relationship between words, is adopted to measure the similarity. As the business category includes multiple categories, the mean value of the categories is adopted to measure the similarity by comparing it with the embedding value of the image category.

4. Experiment

4.1. Dataset. As described in the previous section, images of user interest were collected using the frequency of business categories in Yelp business data by selecting the category that appears most frequently and points to the details and using it as a keyword of a web search engine. Figure 5 presents the frequency of business categories in the Yelp data. The categories with the highest frequencies are restaurants, food, and shopping. However, these categories were not selected as the user's interest category because the range of products or services is wide; instead, categories such as "American food," "boating," and "fitness" were selected as the user's interest category and adopted as web crawling keywords to build the user's interest image dataset. The number of categories in the dataset is 10, and

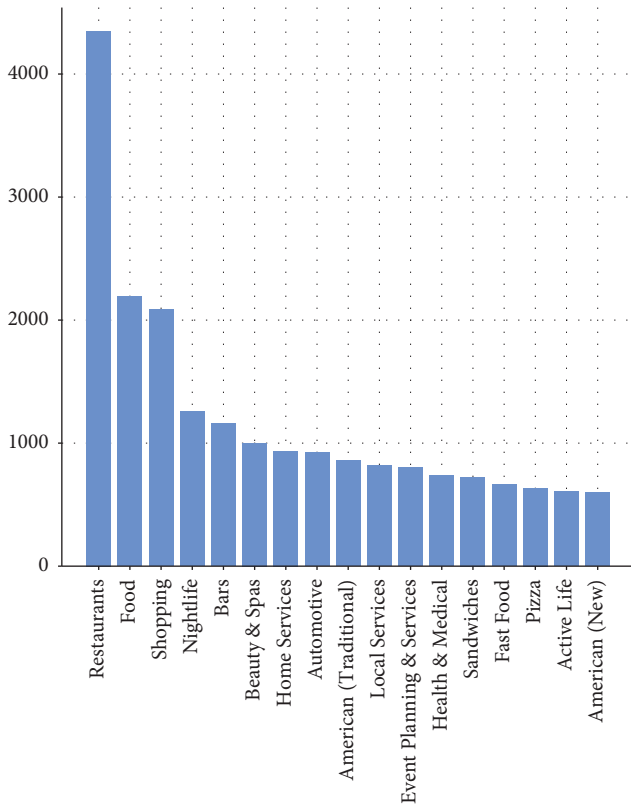


FIGURE 5: Frequency of business categories in Yelp data.

the number of images in each category is presented in Figure 6. The original images were augmented twice to create 40,540 images in the dataset, in which the images were divided at a ratio of 6:3:1 for train, validation, and test data to train the deep learning models.

Table 1 presents Yelp business data, which is the business information data. The Yelp business dataset adopted in this study is a dataset comprising store categories and user reviews, and the names and categories of the businesses, as well as location information such as state, latitude, and longitude were used in this study.

5. Results

5.1. Training Accuracy and Loss of Image Classification. The computer used for the experiment was equipped with Intel i7-9750H and NVIDIA GeForce RTX 2080 Ti model. In total, 20 epochs were applied for the experiment of all three models, 50 layers were fixed and trained for the ResNet and Inception v3 models, while the entire layers were all trained for the EfficientNet B0 model. Figure 7 presents the results for training ResNet, Inception v3, and EfficientNet B0 with 10 category images. As illustrated in Figures 7(a) and 7(c), the ResNet and Inception v3 models exhibited improved accuracies in the training data; however, the accuracy in the validation data did not improve after 70%, as the number of training increased. Figure 7(e) presents the accuracy of the EfficientNet B0 model. Unlike the ResNet and Inception v3 models, the EfficientNet B0 model gradually improved the accuracy in both training

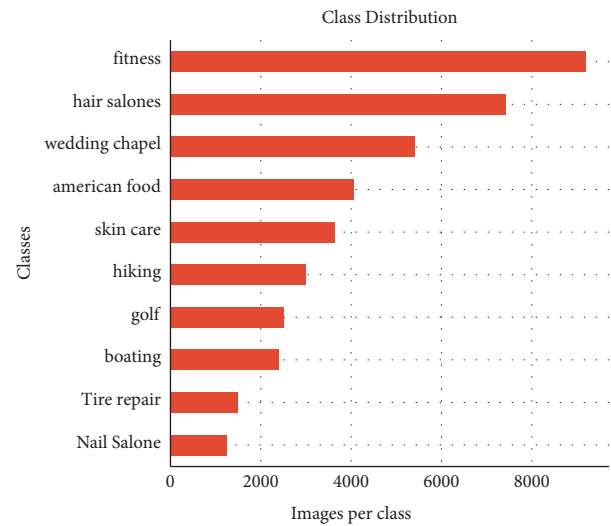


FIGURE 6: Data distribution.

and validation data as training proceeded. Therefore, the EfficientNet B0 model can be considered appropriate for image classification in this study. Figure 8 illustrates the comparison among accuracies of the three models when trained for 20 epochs. At 20 epochs, the ResNet model exhibited 96.71% and 75.74% accuracies in the training and validation data, respectively. The Inception v3 model exhibited 95.32% and 74.3% accuracies in the training and validation data, respectively. In addition, the EfficientNet B0 model exhibited 88.44% and 85.31% accuracies in the training and validation data, respectively.

5.2. Confusion Matrix of Image Classification. A confusion matrix is a measure used to evaluate the performance of deep learning classification. In binary classification, a confusion matrix represents the instances of true positive, true negative, false positive, and false negative. In general, higher true positive and true negative values and lower false positive and false negative values indicate good performance. A confusion matrix in multiclass classification is an excellent performance indicator, as the prediction of each class can be identified. Figure 9 presents the confusion matrix created based on the prediction results of the ResNet, Inception v3, and EfficientNet B0 models after training, where the maximum value is 1. The confusion matrix in Figure 10 demonstrates that the value in the diagonal direction is true positive, where the accuracy is higher as the value is closer to 1. When the test data were applied to each model, the ResNet, Inception v3, and EfficientNet B0 models exhibited 76%, 74%, and 85% accuracies, respectively. According to the confusion matrix, the prediction accuracy was low in the category related to nail salons, which can be improved if more data are collected and refined.

5.3. Category Matching Using Word2vec. Although it is important to accurately classify a user's input image into the proper category, it is also crucial to select candidates for business recommendations for users based on the classification

TABLE 1: Example of Yelp dataset.

Names	Categories	State	City	Address	Latitude	Longitude
Middle East Deli	Food, restaurants, grocery	NC	Charlotte	4508E Independence Blvd	35.194894	-80.767442
LA fitness	Yoga, fitness and instruction, gyms	AZ	Phoenix	7640W Thomas Rd	33.481551	-112.223225
Bank of America	Banks, financial services	NC	Charlotte	9715 Callabridge Ct	35.3275006	-80.9440573
Starbucks	Food, Coffee, and tea	AZ	Phoenix	1610N. 75Th Avenue, #100	33.466005	-112.221288

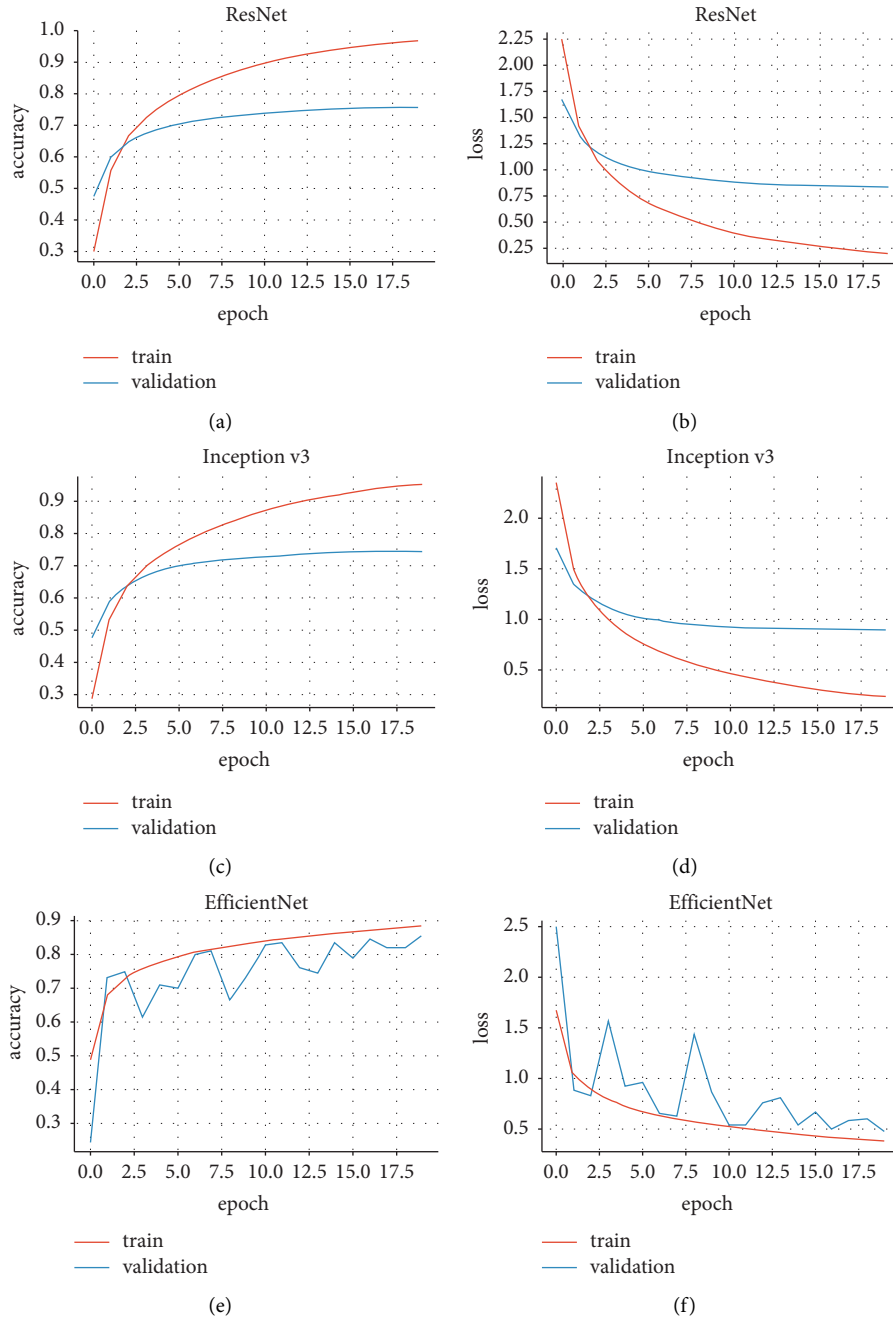


FIGURE 7: Training accuracy and loss of each model. (a) ResNet train accuracy. (b) ResNet train loss. (c) Inception v3 train accuracy. (d) Inception v3train loss. (e) EfficientNet B0 train accuracy. (f) EfficientNet B0 train loss.

results. Therefore, in this section, multiple business categories were applied using Word2vec when measuring the similarity between image and business categories.

Figure 10 presents the method for measuring the similarity between image and business categories using Word2vec. Word2vec learns the relationship between

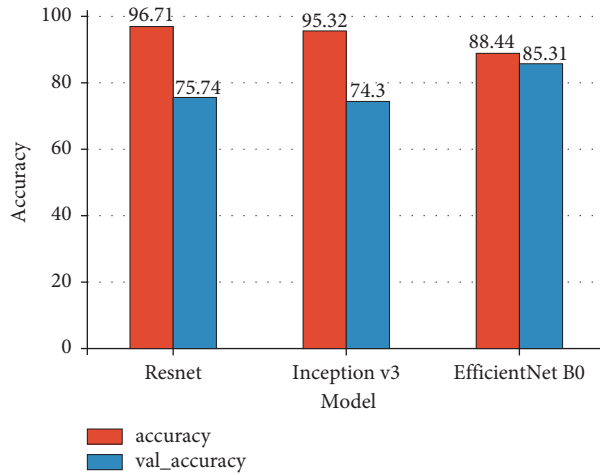


FIGURE 8: Accuracy of each model.

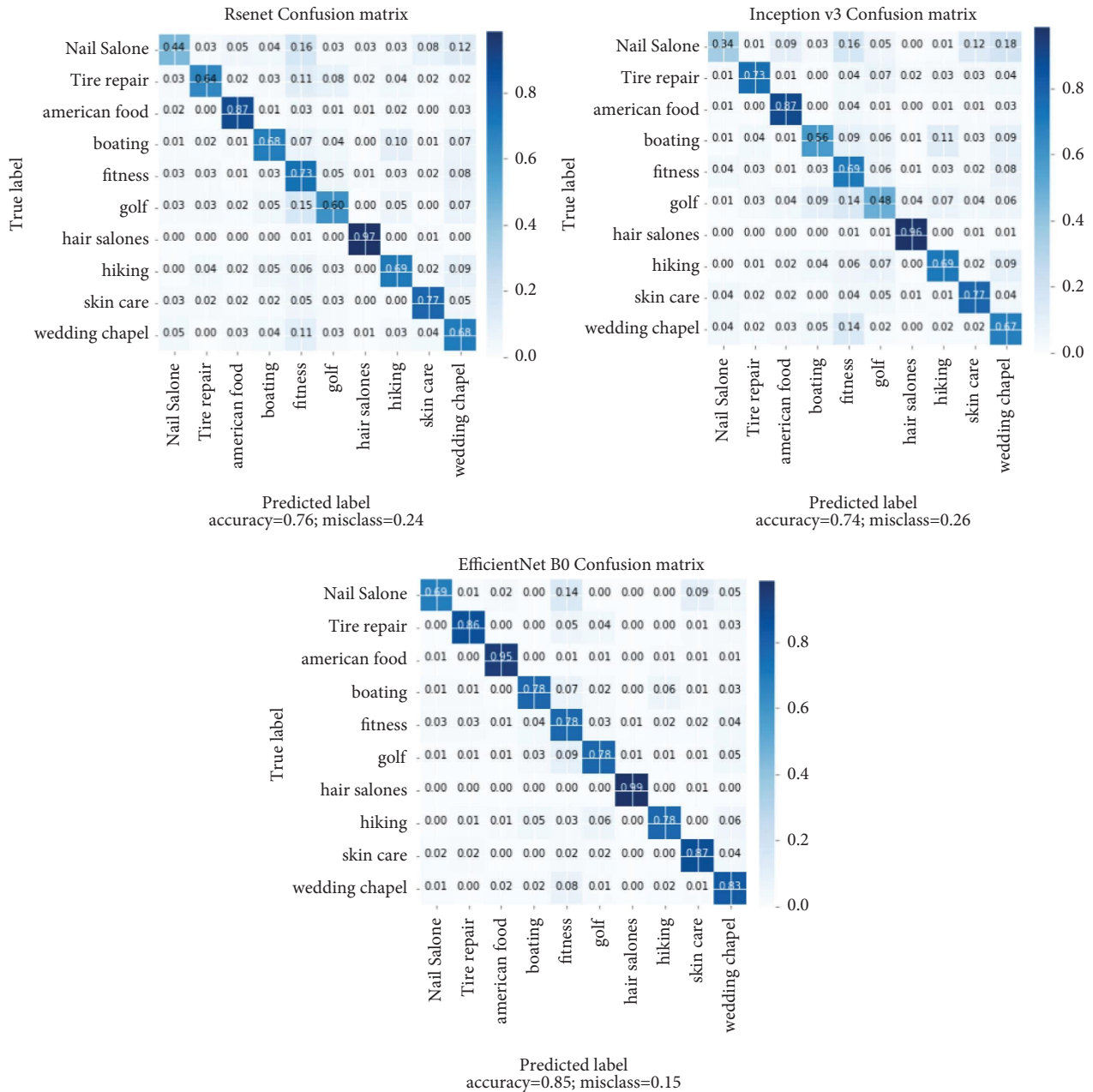


FIGURE 9: Confusion matrix of each model.

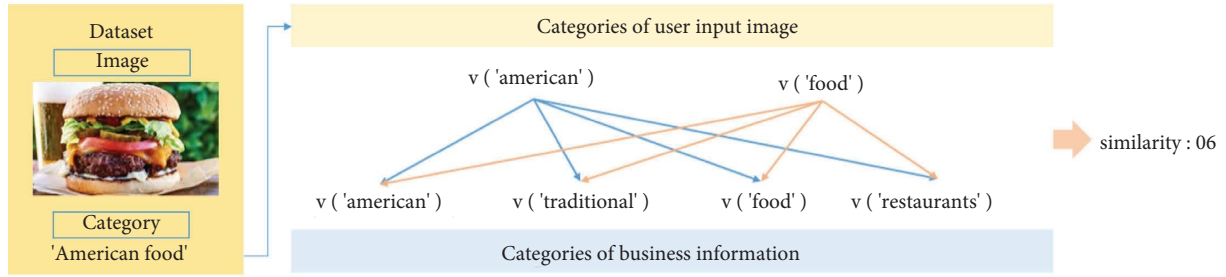


FIGURE 10: Measurement similarity between images and business categories using Word2vec.

TABLE 2: Measurement similarity between “American food” and other business categories.

Image categories	Business categories	Similarity
“American,” “food”	“American,” “traditional,” “food,” “Cajun,” “creole,” “fast,” “food,” “restaurants,” “chicken,” “wings”	0.6
	“Paddleboarding,” “boating,” “rafting,” “kayaking,” “jet,” “skis,” “active,” “life”	0.14
	“Active,” “life,” “trainers,” “gyms,” “fitness,” “instruction”	0.15
	“Professional,” “services,” “local,” “services,” “carpet,” “cleaning”	0.24
	“Parks,” “active,” “life,” “hiking”	0.14
	“Eyelash,” “service,” “beauty,” “spas,” “nail,” “salons”	0.17
	“Eyebrow,” “services,” “hair,” “salons,” “beauty,” “spas”	0.18
	“Automotive,” “tires,” “oil,” “change,” “stations,” “transmission,” “repair,” “auto,” “repair”	0.19
	“Wedding,” “planning,” “Italian,” “caterers,” “party,” “event,” “planning,” “restaurants,” “venues,” “event,” “spaces,” “event,” “planning,” “services”	0.19
	“Country,” “clubs,” “arts,” “entertainment,” “golf,” “active,” “life”	0.25

TABLE 3: Measurement similarity between “hair salon” and other business categories.

Image categories	Business categories	Similarity
“Hair,” “salon”	“American,” “traditional,” “food,” “Cajun,” “creole,” “fast,” “food,” “restaurants,” “chicken,” “wings”	0.17
	“Paddleboarding,” “boating,” “rafting,” “kayaking,” “jet,” “skis,” “active,” “life”	0.13
	“Active,” “life,” “trainers,” “gyms,” “fitness,” “instruction”	0.13
	“Professional,” “services,” “local,” “services,” “carpet,” “cleaning”	0.19
	“Parks,” “active,” “life,” “hiking”	0.06
	“Eyelash,” “service,” “beauty,” “spas,” “nail,” “salons”	0.35
	“Eyebrow,” “services,” “hair,” “salons,” “beauty,” “spas”	0.48
	“Automotive,” “tires,” “oil,” “change,” “stations,” “transmission,” “repair,” “auto,” “repair”	0.07
	“Wedding,” “planning,” “Italian,” “caterers,” “party,” “event,” “planning,” “restaurants,” “venues,” “event,” “spaces,” “event,” “planning,” “services”	0.16
	“Country,” “clubs,” “arts,” “entertainment,” “golf,” “active,” “life”	0.13

words and converts words into numbers. Therefore, the “tire repair” category was separated into “tire” and “repair” via preprocessing, while the business category was also divided into individual words for word embedding. The similarity comparison between image and business categories is the N : N comparison when categories are divided into words; therefore, the average cosine similarity between the image and business-category words was adopted as the similarity between the respective image and category. A model pretrained with Google news data were adopted as the Word2vec model, and the data of the state of Ohio in the Yelp local business data were utilized. Tables 2 and 3 present a part of the similarity measurement results.

Tables 2 and 3 present the results obtained from measuring the similarity between image classification categories “American food” and “hair salon” with other business categories, respectively. In Table 2, the businesses in the first row are most similar to “American food,” while low scores were recorded in other business types. In Table 3, the businesses in the seventh row are the most similar business category to “hair salon.” The business categories in the sixth row exhibited the second highest similarity, which is also illustrated in Figure 11 where “hair,” “nail,” and “skin” have similar values after word embedding, possible owing to the presence of a similar word “beauty” in the category. In the results obtained from measuring the similarity between image and business categories using Word2vec, the business

relevant images were collected via web crawling. Then, in the first stage, different deep learning-based image classification models were trained and tested on the acquired data. The results indicated that the EfficientNet B0 model exhibited better performance on the test data than its counterparts. In the second stage, the similarity between image and business categories was measured using Word2vec and converted to vectors; then, the cosine similarity was measured between each word to adopt the sum of the average values as the degree of similarity. Consequently, the business category that was visibly similar to the image category also exhibited a high degree of similarity based on Word2vec. In addition, the proposed recommendation system can be used for different scenarios, such as, when providing recommendation services to passengers and drivers searching for nearby businesses using the intelligent vehicle infotainment system. In the future, we plan to improve the accuracy of the classifier in the early steps as it may improve the overall performance of the proposed recommendation system and compare it with other classification models.

Data Availability

YELP dataset used to support the findings of this study are available at <https://www.yelp.com/dataset>.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This research was supported by the MSIT (Ministry of Science, ICT), Korea, under the High-Potential Individuals Global Training Program (Grant no. 2020-0-01578) supervised by the IITP (Institute for Information & Communications Technology Planning & Evaluation). This work was also supported by the National Research Foundation of Korea (NRF) grant funded by the Korea Government (MSIT) (Grant no. NRF-2020R1A2C2007091).

References

- [1] J. E. Siegel, D. C. Erb, and S. E. Sarma, "A survey of the connected vehicle landscape—architectures, enabling technologies, applications, and development areas," *IEEE Transactions on Intelligent Transportation Systems*, vol. 19, no. 8, pp. 2391–2406, 2018.
- [2] R. Coppola and M. Morisio, "Connected car: technologies, issues, future trends," *ACM Computing Surveys*, vol. 49, no. 3, pp. 1–36, 2016.
- [3] J. Guo, B. Song, Y. He, F. R. Yu, and M. Sookhak, "A survey on compressed sensing in vehicular infotainment systems," *IEEE Communications Surveys & Tutorials*, vol. 19, no. 4, pp. 2662–2680, 2017.
- [4] G. Sinnapolu and S. M. Alawneh, "Intelligent wearable heart rate sensor implementation for in-vehicle infotainment and assistance," *Internet of Things*, vol. 12, Article ID 100277, 2020.
- [5] L. Wang, H. Yang, X. Qi, J. Xu, and K. Wu, "ICast: fine-grained wireless video streaming over Internet of intelligent vehicles," *IEEE Internet of Things Journal*, vol. 6, no. 1, pp. 111–123, 2019.
- [6] G. J. Dimitrakopoulos and I. E. Panagiotopoulos, "In-vehicle infotainment systems: using bayesian networks to model cognitive selection of music genres," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 11, pp. 6900–6909, 2021.
- [7] S. Hussain, U. Mahmud, and S. Yang, "Car e-talk: an IoT-enabled cloud-assisted smart fleet maintenance system," *IEEE Internet of Things Journal*, vol. 8, no. 12, pp. 9484–9494, 2021.
- [8] A. L. Kun, S. Boll, and A. Schmidt, "Shifting gears: user interfaces in the age of autonomous driving," *IEEE Pervasive Computing*, vol. 15, no. 1, pp. 32–38, 2016.
- [9] Y. Zheng, Y. Liu, and J. H. Hansen, "Intent detection and semantic parsing for navigation dialogue language processing," in *Proceedings of the 2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC)*, pp. 1–6, Yokohama, Japan, October, 2017.
- [10] C. Xu, Z. Guan, W. Zhao et al., "Recommendation by users' multi-modal preferences for smart city applications," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 6, pp. 4197–4205, 2021.
- [11] Apple, "Apple CarPlay," <http://www.apple.com/ios/carplay> (accessed on 2021).
- [12] Google, "Android AU," <https://www.android.com/%20auto/>.
- [13] T. Ranzenberger, C. Hacker, and F. Gallwitz, "Integration of a Kaldi speech recognizer into a speech dialog system for automotive infotainment applications," *Studientexte zur Sprachkommunikation: Elektronische Sprachsignalverarbeitung*, vol. 2018, pp. 293–300, 2018.
- [14] F. Parada-Loira, E. González-Agulla, and J. L. Alba-Castro, "Hand gestures to control infotainment equipment in cars," in *Proceedings of the 2014 IEEE Intelligent Vehicles Symposium*, pp. 1–6, Dearborn, MI, USA, June, 2014.
- [15] B. Verma and A. Choudhary, "Framework for dynamic hand gesture recognition using Grassmann manifold for intelligent vehicles," *IET Intelligent Transport Systems*, vol. 12, no. 7, pp. 721–729, 2018.
- [16] J. Bobadilla, S. Alonso, and A. Hernando, "Deep learning architecture for collaborative filtering recommender systems," *Applied Sciences*, vol. 10, no. 7, p. 2441, 2020.
- [17] B. Zhang, M. Zhu, M. Yu, D. Pu, and G. Feng, "Extreme residual connected convolution-based collaborative filtering for document context-aware rating prediction," *IEEE Access*, vol. 8, pp. 53604–53613, 2020.
- [18] Y. Guan, Q. Wei, and G. Chen, "Deep learning based personalized recommendation with multi-view information integration," *Decision Support Systems*, vol. 118, pp. 58–69, 2019.
- [19] G. Gupta and R. Katarya, "Research on understanding the effect of deep learning on user preferences," *Arabian Journal for Science and Engineering*, vol. 46, no. 4, pp. 3247–3286, 2020.
- [20] S. Zhang, L. Yao, A. Sun, and Y. Tay, "Deep learning based recommender system: a survey and new perspectives," *ACM Computing Surveys*, vol. 52, no. 1, pp. 1–38, 2019.

- [21] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770–778, Las Vegas, NV, USA, June, 2016.
- [22] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, “Rethinking the inception architecture for computer vision,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2818–2826, Las Vegas, NV, USA, June, 2016.
- [23] M. Tan and Q. Le, “Efficientnet: rethinking model scaling for convolutional neural networks,” in *Proceedings of the International Conference on Machine Learning PMLR*, Baltimore, MD, USA, July, 2019.