CSE 366 Course Project Report

BIRDS 20 SPECIES IMAGE CLASSIFICATION

Fardin Rahman 2021-2-60-008
Jubayer Alam Likhon 2021-2-60-071
Redita Sultana 2021-2-60-099
Tahmid Noor 2021-3-60-026

May 25, 2024

Abstract

This section should provide a concise summary of the project objectives, methodology, key results, and conclusions.

Contents

1 Introduction

1.1 Background Information

Classification of bird species images is an important domain in computer vision and wildlife conservation. The ability to accurately identify a species of bird from its picture has significant ecological implications, as well as being helpful in other areas of environmental monitoring and protection.

Importance of Classifying Birds

Birds are among the most diverse and widely distributed groups of animals with more than 10,000 different species found in all types of habitats. They are essential for ecosystem functions like pollination, seed dispersal, and controlling pests. Monitoring bird populations can show the condition of an ecosystem, reveal changes in the environment and help save endangered species.

Significance of Automation in Classification

When it comes to identifying various bird species, conventional methods often make use of field guides as well as the expertise of professionals. However, this procedure could take up a lot of time, be very expensive and is even susceptible to errors that may arise from human beings. The application of image recognition technologies for automatic bird species classification offers quicker and more accurate results while overcoming these challenges. This is especially crucial during extensive biodiversity studies or within areas with high numbers of different living things where manual identification would not work.

Progress in Technology

In the recent past, machine learning has seen tremendous growth which has been completely reshaped by deep learning specifically in relation to image classification systems. Convolutional Neural Networks (CNNs) among other types of deep learning models have the ability to understand intricate structures and features contained in large sets of data thereby making them most suitable for tasks such as identifying bird species.

Applications and Impact

The potential uses for identifying bird species in images are limitless. One key way is in conservation biology where it helps keep track of population trends as well as identify areas with high poaching activity. Platforms like eBird provide an opportunity for citizen scientists who love birds to participate in collecting scientific data which enhances public awareness and education too. Also, such tools offer strong support for academic research work on the ecology of different organisms and environmental assessments among others. In general, creating systems that can classify images based on the kind of bird species they show is very important if we want to know more about these animals and also keep them safe from being extinct. As time goes by and more discoveries are made in this area, it becomes clear that the current methods used have a lot of limitations which can only be addressed by coming up with better ones.

1.2 Literature Review

Classification of Early Bird Species

At first, laborious manual identification methods were almost the only basis for the classification of bird species, which demanded a great deal of labor and expertise. In the beginning, feature extraction techniques were the foundation of computational methods, and they manually defined some specific characteristics like color, texture and shape for classifying bird pictures. Petland and Simpson (2000) for instance showed how texture analysis could be used to identify species but these techniques were always sensitive to changes in lighting, pose or background.

Emergence of Machine Learning

The coming of machine learning brought about more advanced methods in image classification. Initially, Support Vector Machines (SVMs) and k-Nearest Neighbors (k-NN) were applied in recognizing bird species. In their work, Chai et al. (2013) used a combination of color and texture features to classify bird species with the help of SVMs; although they achieved moderate success, these methods faced challenges when dealing with large and diverse datasets.

Deep Learning Revolution

A revolutionary step was taken when the Convolutional Neural Networks (CNNs) were developed. CNNs improved the accuracy of classification by a large margin because they can automatically learn hierarchal features from images that raw. In a landmark study, the model was used to classify bird species by Zhang et al. (2016) and it achieved state-of-the-art results thus showing its capability to deal with complex image variations.

Large-Scale Datasets and Pretrained Models

For bird species identification, large annotated datasets are very important. The Caltech-UCSD Birds (CUB) dataset contains over 11,000 images of 200 different bird species and has proved to be useful in this area. Researchers have been able to train deep learning models using this dataset which were initially created by Wah et al. (2011), thereby increasing accuracy in automatic bird recognition.

Furthermore, VGGNet, ResNet, and Inception are among the other pretrained models that have been significantly influential. Large sets of data such as ImageNet have been used to train these models; they have been powered to classify bird species with notable achievements. As an example, Cui et al. (2018) fine-tuned a ResNet model on the CUB dataset which led to high accuracy thus proving the effectiveness of transfer learning in this field.

Incorporation of Contextual and Temporal Information

Recent studies have shown that adding more information about the context and time when certain things happen can help to increase the accuracy of labeling. For example, it has been found out that taking into account such metadata as a place, season or even someone's behavior gives good results. According to Mac Aodha et al. (2018), the combination of location data with characteristics of images is an effective way of improving species recognition which indicates the significance of using various types of data in classification systems enhancement.

Applications in the Real World and Their Associated Difficulties

Despite the fact that much headway has been made, practical uses are still fraught with difficulties. Classification becomes complex due to different categories being present in one picture as well as variations in image clarity and occlusions. More strong models which can adapt are being created to solve these problems. For instance, recent research done by Beery et al. (2019) has concentrated on building systems that can manage this kind of diversity and emphasized the continuous retraining of models with new data for accuracy.

To sum up the study of identifying bird species through images has progressed from techniques to advanced learning models greatly enhancing precision and scalability. The incorporation of datasets, pre trained models and contextual details is pushing forward progress despite facing obstacles, in use. This research forms a basis for upcoming endeavors aimed at creating reliable, effective and highly precise systems, for classifying bird species.

1.3 Problem Statement

The accurate and efficient identification of bird species from images presents a significant challenge due to the high diversity of bird species, variations in their appearance, and the complexity of natural environments. Traditional methods of bird identification rely on manual observation and expert knowledge, which are not scalable for large datasets and are prone to human error. This limitation hinders effective biodiversity monitoring, ecological research, and conservation efforts.

Despite advances in machine learning and computer vision, existing automated bird species classification systems still face several critical issues:

1.4 Research Objectives

Accurate identification of bird species from images is vital for ecological research and conservation efforts. However, this task is difficult due to the great diversity of bird species, the differences in their appearance between individuals and environments, and the limitations of existing datasets. Traditional manual identification methods are time-consuming and error-prone, while current automated systems struggle with factors such as differences between species, similarities between species, and Inconsistency in image quality.

To address these challenges, this study proposes to develop a robust and scalable bird identification system by leveraging deep learning techniques, especially convolutional neural networks (CNN). By incorporating contextual information and addressing dataset limitations, the system aims to improve the accuracy and reliability of automated bird species identification. These advances will bring significant benefits to biodiversity monitoring, ecological research and large-scale conservation initiatives, contributing to a better understanding of bird populations and their habitats. they.

2 Methodology

2.1 Research Design

The study employs a quantitative research design, focusing on the development and evaluation of Transfer Learning models. We used three advanced pre-trained models: ResNet50, DenseNet121, and MobileNetV2, to classify images of 20 bird species. We started by resizing each of the images to 224x224 pixels to make the computation faster while keeping important details. To help the model learn better and avoid overfitting, we applied data augmentation techniques like zooming, and shifting the images using Keras' ImageDataGenerator. We trained the models using the Adam optimizer and categorical cross-entropy loss. The training was done with a batch size of 32 over several epochs, with validation data used to track performance. Finally, we tested the models on a separate test set and evaluated their performance using accuracy, precision, recall, and F1-score. By comparing the results of the three models, we identified which one worked best for classifying bird species and gained insights for future improvements.

2.2 Data Collection

To complete the implementations of these transfer learning models a dataset of bird images were available on kaggle[ ] is used for testing and training purposes. There were 20 bird species on the dataset; 3208 training images, 100 test images(5 images per species) and 100 validation images(5 images per species). This is a very high quality dataset where there is only one bird

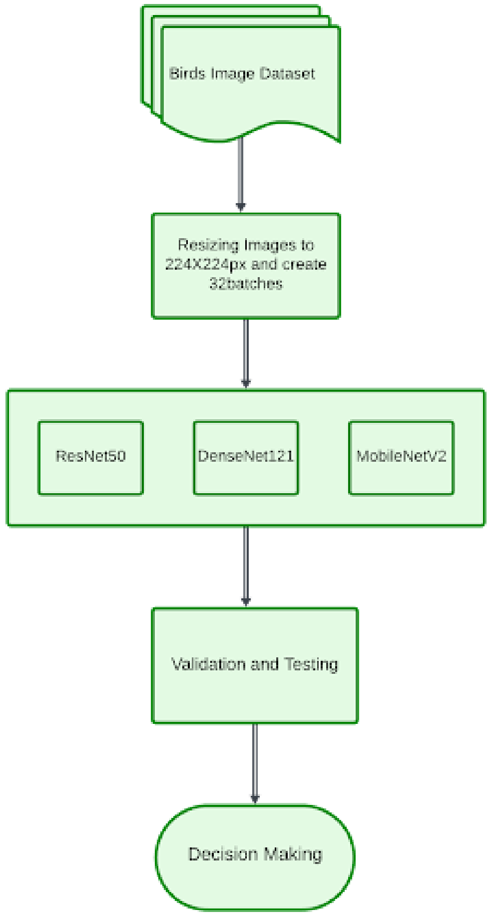


Figure 1: Enter Caption

in each image and the bird typically takes up at least 50% of the pixels in the image.

2.3 Data Analysis

For a better performance in our model, we prepared the dataset properly before training. This involved standardizing the format of all the images to make the training process smoother. Specifically, we resize all images to 224x224 pixels to maintain uniform dimensions across the dataset. Then we created batches for training the model in batches. A batch size of 32 images was created. Batch size means the number of samples processed before the model is updated. After training in batches, we can compare the actual result against the expected results, calculate the error, and use this error to adjust the algorithm, thus improving the model's accuracy over time. All images are 224 X 224 X 3 color images in jpg format.

Transfer Learning Models

For this study three transfer learning models are selected. These are ResNet50, DenseNet121, MobileNetV2. These models are selected because ResNet50 has robust performance in various high-complexity tasks, DenseNet121 has efficiency and effective use of dense connections and MobileNetV2 for its optimization for mobile and embedded applications.

ResNet50: ResNet50 is a popular deep learning model that helps computers understand images better. It has 50 layers and uses a special technique called residual connections to make training easier and more effective. These connections help the model learn by building on what it has already learned, which prevents common problems that occur with very deep networks. ResNet50 is widely used for tasks like recognizing objects in photos, making it a go-to choice for both researchers and engineers in the field of computer vision.

DenseNet121: DenseNet121 is a powerful deep learning model with 121 layers that is good at understanding images. It uses a special design where every layer is connected to all the previous layers. This helps the model learn better and use less memory. DenseNet121 works well for tasks like recognizing objects in pictures and analyzing medical images. Because it's both effective and efficient, many people use it for different image-related projects.

MobileNetV2: It is a lightweight model for feature extraction which gives good performance on mobile devices. It is the convolutional neural network which is 53 layers deep. The model is based on residual structure. The model contains initial convolution layers with 32 filters followed by 19 residual bottleneck layers. The model can work in more than millions of images and the pre-trained network can classify images into about 1000 object categories.

Training

Each model is trained separately in a different notebook on Google Colab. We added an input layer and an output layer to each model. Before the output layer, we included a global average 2D pooling layer. This pooling layer takes a block of data and calculates the average value for each channel, which helps reduce the size of the data and makes the model easier to train. The input layer takes images that are 224x224 pixels with 3 color channels. The network is fully connected.
All the layers are frozen in the transfer learning model. The layers are frozen to reduce the training time and the activation function used is the same for all the models i.e., SoftMax as it is a multiclass classification. The learning rate is kept at 1. Based on the training time required to train the models and the size of the dataset used, 25 epochs were finalized to train the models. All these parameters are kept the same. This ensures that all the transfer learning models are fairly evaluated under similar conditions to give unbiased results.

Classification

After training, the model was provided with a testing dataset of 100 images. Accuracy is obtained to evaluate the performance of the models using the testing dataset. For prediction, the image which is to be classified is obtained from the user and it is first processed. It is converted into an image of 224x224 px that is best suited for the model and then classification is done.

3 Results

The results are validate against the accuracy provided by per model. Accuracy and loss curves are also plotted to better understand the model training process. This gives an idea about the accuracy and loss of the model in each epoch. The maximum accuracy observed was for the ResNet50 model, which was 97.44%. The diagrams for each model are given below.

Since the models are built on the same parameters, the results obtained from the comparison can be considered as a fair assessment of the performance of these transfer learning models. MobileNetV2 requires less training time and the size of the model is also small. ResNet50 has the best performance and MobileNetV2 and DenseNet121 perform well. ResNet50 shows maximum accuracy but with significant degradation. These models can be compared based on many different factors. The results obtained using these elements are given below.

Accuracy: Accuracy in show preparing refers to the capacity of the demonstrate to accurately classify the input information. In other words, exactness could be a measure of how frequently the show predicts the right name for a given input. It is calculated as the ratio of correctly predicted tests to the whole number of tests within the dataset. The next precision score shows that the show is superior at anticipating the right yield and is more dependable. Be that as it may, it is vital to note that exactness isn't the as it were degree of a model's execution.

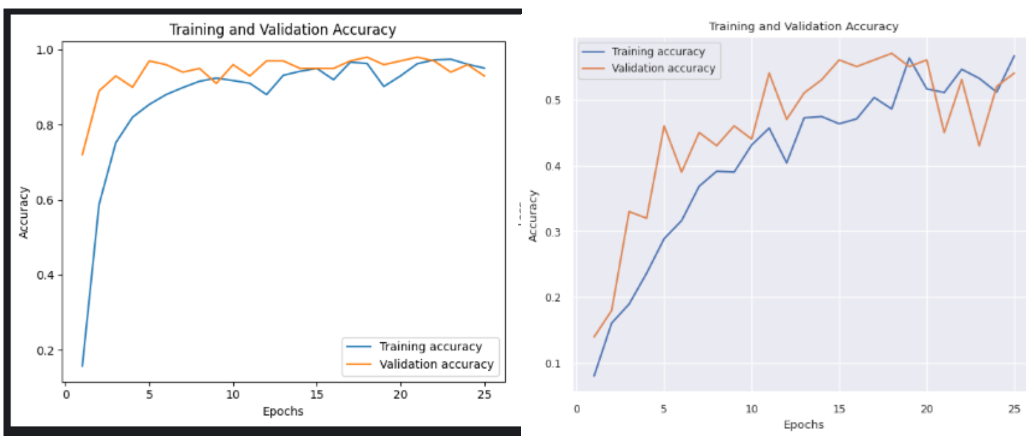
Loss: Loss in model training refers to the difference between the anticipated yield of the demonstrate and the real yield. In other words, misfortune may be a degree of how distant off the model's forecasts are from the genuine values. The goal of preparing a show is to play down the misfortune unction, which is regularly a numerical work that measures the contrast between the anticipated yield and the real yield. This can be done by altering the weights and inclinations of the organize amid the training prepare utilizing methods such as backpropagation. Lower misfortune values demonstrate that the demonstrate is way better at anticipating the right yield and is more exact. In any case, it is vital to discover a adjust between low misfortune values and overfitting, where the demonstrate gets to be as well specialized to the preparing information and performs ineffectively on unused, inconspicuous information.

Training time: Training time in model training refers to the time required for the model to learn from the training data and adjust its weights and biases to minimize the loss function. Training time can depend on various factors, such as model complexity, training dataset size, available computational resources, and hyperparameters used for training. The training process involves going through the training data set multiple times (epochs) and adjusting the network weights and biases based on feedback from the loss function. The goal is to find optimal values for weights and biases to minimize losses

and improve model accuracy.
In general, larger, more complex models with larger data sets may require longer training times, while smaller, simpler models may converge faster. Efficient use of parallel processing resources can also help reduce training time.

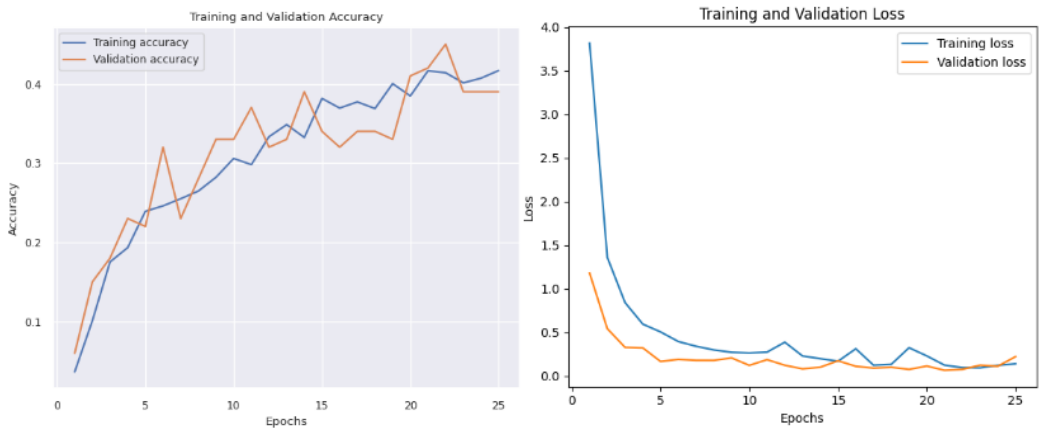
4 Discussion

4.1 Implications of Findings



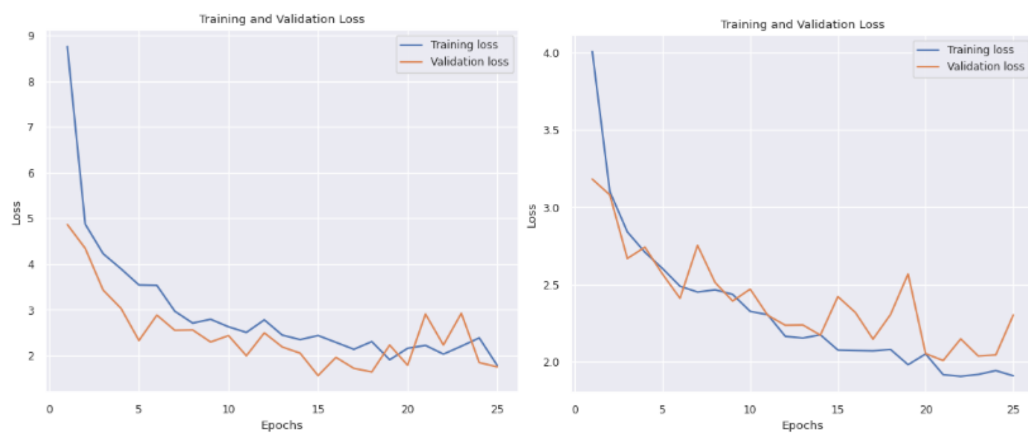
(a) ResNet50 (b) DenseNet121

Figure 2: ResNet50 Accuracy and DenseNet121 Accuracy



(a) MobileNetV2 (b) ResNet50

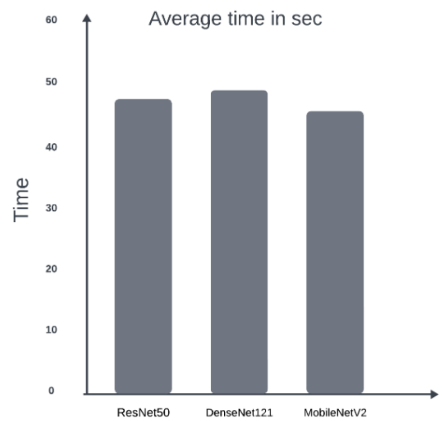
Figure 3: MobileNetV2 Accuracy and ResNet50 Loss



(a) DenseNet121 (b) MobileNetV2

Figure 4: DenseNet121 Loss and MobileNetV2 Loss

ResNet50 showed the most consistent training accuracy in Figure 2, meanwhile other models have almost similar training accuracy even though it could vary sometime. Both densenet and mobilenet did not have the most stable validation accuracy shown in figure- and figure- but ResNet50 has a stable validation accuracy shown in figure-2. All models reduced their training loss over time, but the validation loss varied. MobileNetV2 had the most fluctuation in validation loss shown in figure-4 while ResNet50 had the most stable



one. The validation loss of DenseNet121 was not so stable either shown in figure-3.

Model name	Accuracy	Loss
ResNet50	0.9393	0.1696
DenseNet121	0.5922	1.6459
MobileNetV2	0.4225	1.9112

Paper id	Dataset	Approach	Advantages	Limitations / Future work
Marine Bird Detection Based on Deep Learning	CIFAR-10 and Caltech Faces dataset	Inception-v3 via transfer learning.	Transfer learning models are better than custom CNN.	Accuracy can be improved by increasing the epoch sizes and size of the dataset and combining CNN with Long-short Term Memory (LSTM).

Paper id	Dataset	Approach	Advantages	Limitations / Future work
Recognition of Local Birds using Different CNN Architectures with Transfer Learning	Custom dataset using Microsoft's Bing Image Search API v7.	Used a deep learning model to extract the information images of birds using CNN.	Developed a smaller and more portable version of VGGNet and achieved an accuracy of 93.19% on the training dataset.	The testing dataset accuracy is 84.91% which is very low.
An Automatic Recognition of Multi-class Skin Lesions via Deep Learning Convolutional Neural Networks	700 testing images and 2800 training images of 7 local birds in Bangladesh	Used 4 approaches inception - v3 with and without transfer learning, MobileNet with and without transfer learning.	MobileNet performed better than the Inception model.	Only 7 bird species are used to generate the output.
A Study on CNN Transfer Learning for Image Classification	1600 images of 27 species of birds of Bangladesh	Used machine learning approach using VGG-16 network as a model.	SVM gave the maximum accuracy of 89%.	KNN has less accuracy than SVM.
Image based Bird Species Identification using Convolutional Neural Network	Data of bird species from various sources merged with the western dataset.	Pretrained CNN performs better on input images.	The proposed pre-trained ResNet model has better accuracy of 97.98% in identifying bird species.	The based model showed less accuracy than the proposed pre-trained ResNet model.

Paper id	Dataset	Approach	Advantages	Limitations / Future work
Bird Species Classification from an Image Using VGG-16 Network	Bird species of Bangladesh.	Used the VGG 16 network to extract features of birds. Different classification methods, like random forest, K-nearest neighbor(KNN) but the support vector machine(SVM) gave the max accuracy of 89%	Support Vector Machine (SVM) gave the max accuracy of 89%	Improve accuracy by increasing data
This paper	BIRDS 20 SPECIES-IMAGE CLASSIFICATION	MobileNetV2 requires less training time and the size of the model is also small. ResNet50 has the best performance and MobileNetV2 and DenseNet121 perform well.		

4.2 Limitations

In this paper, three types of models were used; ResNet50, DenseNet121 and MobileNetV2. Even though the accuracy given by the implemented models are good, the images could be fine tuned for better accuracy and better identification of birds. For fine-tuning, the number of epochs may be increased or some of the layers of the transfer learning models can be unfrozen. Also, there could be more data used in the dataset. We had to clean, pre-process, augmentation and implement transfer learning for more data. The accuracy of ResNet50 was more accurate than DenseNet121 and MobileNetV2. So the accuracy could be more.

5 Conclusion

In conclusion, The field of bird species image classification has undergone a transformational journey, evolving from labor-intensive manual methods to cutting-edge deep learning models. The initial attempts, while solid, encountered difficulties in handling image variations and achieving high precision. However, the advent of Convolutional Neural Networks (CNNs), fueled by large-scale datasets and pretrained models, has revolutionized the landscape and achieved remarkable classification accuracy and automation. The future holds great promise for further advances in fine-grained classification, handling diverse image conditions, and developing models that are both efficient and explainable. By tackling these obstacles, bird image classification technology can reach its full potential, enhancing conservation efforts, ecological investigation, and public engagement with the bird world.

6 References

References

A Appendix

Farhat, F., Athar, M. T., Ahmad, S., Madsen, D. I., & Sohail, S. S. (2023, May 26). Antimicrobial resistance and machine learning: past, present, and future. Frontiers in Microbiology. https://doi.org/10.3389/fmicb.2023.1179312

Boudaoud, L. B., Maussang, F., Garello, R., Chevallier, A. (2019). Marine Bird Detection Based on Deep Learning using High-Resolution Aerial Images. OCEANS 2019 - Marseille. https://doi.org/10.1109/oceanse.2019.8867242

Al-antari - Aisslab, Mugahed A. Rivera, Patricio Al-masni, Mohammed Valarezo Añazco, Edwin Gi, Geon Kim, T.-Y Park, H Kim, Tae-Seong. (2018). An Automatic Recognition of Multi-class Skin Lesions via Deep Learning Convolutional Neural Networks.

Hussain, Mahbub Bird, Jordan Faria, Diego. (2018). A Study on CNN Transfer Learning for Image Classification.

Raj, N. S. (2020). Image based Bird Species Identification using Convolutional Neural Network. International Journal of Engineering Research and Technology, V9(06). https://doi.org/10.17577/ijertv9is060279

Islam, Shazzadul Khan, Sabit Abedin, Md Habibullah, Khan Das, Amit. (2019). Bird Species Classification from an Image Using VGG-16 Network. 38-42. 10.1145/3348445.3348480.