International Journal of Scientific Research in Computer Science, Engineering and Information Technology

ISSN : 2456-3307 OPEN ACCESS

Available Online at : www.ijsrcseit.com

doi : https://doi.org/10.32628/CSEIT2511110

Deep Learning for Image Classification: Methods, Challenges, and Future Directions

Sanjay Kumar Gorai, Anurag Sarangi, Shekhar Pradhan*

Post Graduate, Department of Computer Science, Kolhan University, Westsingbhum, Chaibasa-833201, Jharkhand, India

ARTICLEINFO

Article History:

Accepted : 11 Jan 2025

Published: 13 Jan 2025

Publication Issue

Volume 11, Issue 1

January-February-2025

Page Number

484-496

ABSTRACT

Image classification has been fundamentally changed by deep learning that has driven unprecedented accuracy and has empowered applications ranging from healthcare to autonomous cars to security. For example, medical imaging has been diagnosed for diseases such as diabetic retinopathy and tumour detection using deep learning models to an excellent degree. Object classification algorithms in autonomous vehicles are responsible for enabling real time navigation and obstacle avoidance. More recently, the advances in image classification have been made possible with recent breakthroughs including Vision Transformers (ViTs) and self-supervised learning models like SimCLR. In this paper, we explore the main methods on which the deep learning-based image classification fundamentally lies, including the convolutional neural networks (CNNs), transfer learning, and attention mechanisms. Finally, it also discusses the field challenges, like the need to large labelled datasets, computational requirements, and interpretability and it provides solutions to overcome them. We conclude with promising future directions including few shots learning, unsupervised learning and the combination of multimodal data and how they will further advance and open up new applications.

Keywords: Deep Learning, Image Classification, Convolutional Neural Networks (CNNs), Transfer Learning, Vision Transformers (ViTs), Data Augmentation, Interpretability, Few-Shot Learning, Self-Supervised Learning, Multimodal Learning, Federated Learning

Related Works:

The rapid progress in deep learning of image classification is due to numerous studies. AlexNet [1] pioneered deep convolutional networks, and revolutionized the field. VGGNet improved on this, as Simonyan and Zisserman [6] did with simplicity and depth. Using ResNet, He et al. [7] solve the issue of vanishing

gradient.

Further, attention mechanisms have helped image classification. In line with this, Dosovitskiy et al. [9] proposed Vision Transformers that surpassed CNNs on corresponding datasets by taking advantage of self-attention. Additionally, hybrid models that combine CNNs and transformers have also appeared [17]. Finally, we evaluate these types of advancements on datasets such as ImageNet and CIFAR and prove that they are useful in multiple situations.

Recent transfer learning techniques including Inception [12] and Mobile Net [13] have made it easy to train on limited data. The zero-shot classification capability of the combination of text and image data has been shown with OpenAI's CLIP model [14]. In fact, these approaches have been validated by benchmarks like ImageNet zero-shot accuracy metric.

Along with regularization techniques like dropout [19] and ingenious data augmentation like Cut Mix [20], similar improvements in generalization have also been achieved. These methods were tested on datasets such as COCO and Pascal VOC. Since there has not been a single application which placed better than an ensemble method in competitions [23], it has been evaluation on cross-validation of these datasets that has often resulted in top results.

Both methods for adversarial robustness [29], fairness [30], and interpretability [27] have received substantial attention in addressing challenges. For instance, adversarial training had been widely verified on synthetic datasets such as FGSM and PGD. Benchmarks like Mini-ImageNet and Omniglot offer a robust testing ground for future research on recent advances in Few-shot learning [33], self-supervised learning [34] and federated learning [36].

Introduction

Image classification is one of the basic problems in the computer vision, which is to classify an input image with a label. Prior work relied on handcrafted features and classical machine learning algorithms, generally falling short, particularly because image variability limits the usefulness of these approaches. However, with deep learning, especially with the emergence of CNNs, image classification has been revolutionized by the end-to-end learning of features and decision boundaries right on the raw pixel data.

1.1 Importance of Image Classification

Image classification is critical in many applications:

- Healthcare: For instance, identifying disease from medical imagery by utilizing CNNs to identify lung cancer or diabetic retinopathy from X-rays or retinal scans [1].
- Autonomous Vehicles: Allowing objects to be recognised for navigation, i.e. pedestrians, vehicles, and traffic signs to allow safe decision-making. Deep learning is integrated in Tesla's Autopilot system for real time object detection and classification so that the navigation is efficient and adaptive [2].
- Security: Facial recognition and surveillance systems used at airports, public spaces to monitor, identify people, and such. For instance, the

deployment of facial recognition systems in China has been very useful in ensuring that suspects are caught 'in real time [3].

- Retail: Product recognition and inventory management as part of building the cashier-less checkout systems in stores such as Amazon Go. These models simplify the shopping experience by classifying items placed in virtual carts [4].

1.2 Overview of Deep Learning in Image Classification

Deep learning model (such as CNN) have demonstrated human level performance in many image classification tasks. On the other datasets like image and recorded by benchmarks of architectures like AlexNet [5], VGG [6], ResNet [7], and EfficientNet [8] we set benchmarks of accuracy. Novel approaches, like Vision Transformers (ViTs) and attention-based models, have continued to push the boundaries of performance [9], beyond CNNs.

Methods in Deep Learning for Image Classification

In this section, the image classification methodologies on how the fields of advancement have been driven are examined.

2.1. Convolutional Neural Networks (CNNs)

Deep learning for image classification is built around CNNs. Each one contains convolutional layers, pooling layers and fully connected layers. Key architectures include:

- AlexNet: First deep CNN to win ILSVRC [5].
- VGGNet: Has a long-known simplicity [6] and deep architecture.
- ResNet: They were introduced skip connections to solve the vanishing gradient problem [7].
- EfficientNet: It is an accurate and computationally efficient approach [8].

2.1.1. Evolution of CNN Architectures

Previous work on CNN architectures emphasizes increasing depth and parameter efficiency. As an example, taking AlexNet to this level further, VGGNet used smaller convolutional filters (3x3) and more layers. But deep challenges were then coming in terms of higher computational cost and overfitting into the depth. To combat overfitting, we used techniques such as early stopping, dropout and data augmentation, while the advent of hardware, in the form of GPUs and TPUs, was useful in handling computational needs.

In other words, residual connections tackled the vanishing gradient problem and instrumental in training networks more than 100 layers deep [7]. By demonstrating this innovation in image competition competitions, it greatly improved convergence and accuracy. Further, Efficient Net tailored this further by simultaneously optimizing depth, width and resolution for more efficient per computational cost and better performance [8]. Consisting of a compound scaling that balances network depth, width and resolution, Efficient Net was introduced.

2.1.2. Applications of CNNs

CNNs have proven to be ubiquitous in medical imaging ranging from tumour detection in MRIs [10], object recognition in autonomous vehicles [2] to satellite image analysis for environmental monitoring [11].

2.2. Transfer Learning

Transfer learning relies on pre trained models to learn in new tasks with little labelled data. Popular pre-trained models include:

- Inception: Based on the research work done by Google for scalable image classification [12].
- MobileNet: For mobile and edge devices [13].
- CLIP: It combines text and image data for zero shot classification [14].

2.2.1. Advantages of Transfer Learning

- Improves performance on small datasets.
- Facilitates domain adaptation [15].

2.2.2. Limitations of Transfer Learning

Transfer learning is not straightforward if the source and target domains are very different and require fine tuning techniques to fill the gap [16].

2.3. Attention Mechanisms and Vision Transformers (ViTs)

Classification, when isolated from all other channels, suffers when attention mechanisms are used, implying they can focus on all the wrong parts of an image to adversely affect classification. Being inspired by transformers in NLP, ViTs model global relationships in images using self-attention mechanisms. They have shown state of the art results on benchmark datasets [9].

2.3.1. Vision Transformers vs. CNNs

CNNs excel at modelling local features whereas ViTs excel at modelling long range dependency. Recent models combining CNNs and attention mechanisms are trying to take the best from both [17].

2.4. Data Augmentation and Regularization

Techniques such as rotation, flipping, cropping and colour jittering are used for augmenting training data to enhance model generalization [18]. On that basis, overfitting [19]can be mitigated by regularization such as dropout, weight decay and batch normalization.

2.4.1. Advanced Data Augmentation Techniques

- CutMix: Functions as a patch combining technique from different images [20].
- Mixup: Linear interpolation of pairs of images generates new training samples [21].

2.5. Ensemble Methods

Machine learning technique ensemble methods is the technique of combining several base models to make a single optimum predicting model [22].

But three problems have been overcome by the Ensembles.

- Statistical Problem –The statistical problem happens when there is more data than the size of the hypothesis space. Therefore, the learning algorithm picks out but one hypothesis even though there are many with equal accuracy on the data! We run a risk of confidence in selected hypothesis that is not accurate on unseen data!
- Computational Problem –The above Computational Problem occurs when we cannot guarantee that the learning algorithm would find the best hypothesis.
- Representational Problem –The Root of the Representational Problem is that the hypothesis space does not contain any good approximation of the target classes.

Techniques used in Ensemble method are bagging, boosting, and stacking

1. Bagging: - BAGGing, or Bootstrap AGGregating. BAGGing gets its name because it combines Bootstrapping and Aggregation to form one ensemble model. Given a sample of data, multiple bootstrapped subsamples are pulled. A Decision Tree is formed on each of the bootstrapped subsamples. After each subsample Decision Tree has been formed, an algorithm is used to aggregate over the Decision Trees to form the most efficient predictor.

The steps of bagging are as follows:

- We have an initial training dataset containing n-number of instances.
- We create a m-number of subsets of data from the training set. We take a subset of N sample points from the initial dataset for each subset. Each subset is taken with replacement. This means that a specific data point can be sampled more than once.
- For each subset of data, we train the corresponding weak learners independently. These models are homogeneous, meaning that they are of the same type.
- Each model makes a prediction.
- The predictions are aggregated into a single prediction. For this, either max voting or averaging is used.

2. Boosting: - Boosting is a technique in which we are combining weak learners with high bias. Boosting aims to produce a model with a lower bias than that of the individual models. Like in bagging, the weak learners are homogeneous. Boosting involves sequentially training weak

learners. Here, each subsequent learner improves the errors of previous learners in the sequence. A sample of data is first taken from the initial dataset. This sample is used to train the first model, and the model makes its prediction. The samples can either be correctly or incorrectly predicted. The samples that are wrongly predicted are reused for training the next model. In this way, subsequent models can improve on the errors of previous models. Unlike bagging, which aggregates prediction results at the end, boosting aggregates the results at each step. They are aggregated using weighted averaging. Weighted averaging involves giving all models different weights depending on their predictive power. In other words, it gives more weight to the model with the highest predictive power. This is because the learner with the highest predictive power is considered the most important.

Boosting works with the following steps:

- We sample m-number of subsets from an initial training dataset.
- Using the first subset, we train the first weak learner.
- We test the trained weak learner using the training data. As a result of the testing, some data points will be incorrectly predicted.
- Each data point with the wrong prediction is sent into the second subset of data, and this subset is updated.
- Using this updated subset, we train and test the second weak learner.
- We continue with the following subset until the total number of subsets is reached.
- We now have the total prediction. The overall prediction has already been aggregated at each step, so there is no need to calculate it.

3. Stacking: -In Stacking we are improving the prediction and accuracy of strong learners. Stacking aims to create a single robust model from multiple heterogeneous strong learners.

Stacking differs from bagging and boosting in that:

- It combines strong learners
- It combines heterogeneous models
- It consists of creating a Metamodel. A metamodel is a model created using a new dataset.

Individual heterogeneous models are trained using an initial dataset. These models make predictions and form a single new dataset using those predictions. This new data set is used to train the metamodel, which makes the final prediction. The prediction is combined using weighted averaging. Because stacking combines strong learners, it can combine bagged or boosted models.

The steps of Stacking are as follows:

1. We use initial training data to train m-number of algorithms.
2. Using the output of each algorithm, we create a new training set.
3. Using the new training set, we create a meta-model algorithm.
4. Using the results of the meta-model, we make the final prediction. The results are combined using weighted averaging.

2.5.1. Practical Applications of Ensembles

In competition, top performance has been achieved using ensemble methods for the ImageNet challenge [23] and Kaggle.

Challenges in Deep Learning for Image Classification

Despite its successes, deep learning for image classification faces several challenges:

3.1. Data Requirements

Generating large labelled datasets can be expensive and time consuming, but deep learning models require them. Generative Adversarial Networks (GANs) and computer simulations have presented synthetic data generation as also a valid option to generate real looking training samples. For instance, in order to augment datasets for facial recognition systems GANs have been used to learn novel facial expressions and poses for increasing model robustness. Like in autonomous vehicle research, various

simulated environments, such as CARLA, are used to generate many different driving scenarios so that models can train on the edge cases without risking real world. These efforts are further augmented by semi-supervised learning which exploits large amounts of unlabelled data using techniques including pseudo labelling and consistency regularization [24].

3.1.1 Overcoming Data Scarcity

To reduce the demand for extensive labelled datasets, we employ techniques such as data augmentation, transfer learning and active learning [25].

3.2. Computational Complexity

Deep models require big computational resource to train. Model pruning, quantization, efficient architectures try to reduce complexity [26].

3.3. Interpretability

We must understand the decision-making process of deep models — and it matters especially in high stakes applications like healthcare. This problem is addressed using explainable AI (XAI) tools such as Grad-CAM and LIME [27].

3.3.1 Importance of Interpretability

Debugging aid and compliance with ethical and legal standards [28] is increased by interpretability.

3.4. Robustness and Generalization

Adversarial attacks and domain shifts have been shown to be very sensitive to deep models. There is much work ongoing on adversarial training and domain adaptation [29].

3.5. Ethical Concerns

Unfair outcomes from training data can be caused by bias. To address these issues [30], transparent data curation and fairness aware algorithms are in order.

3.5.1 Addressing Bias in Models

To achieve equitable outcomes, learning is explored through methods such as re-sampling, fairness aware loss functions and adversarial debiasing [31].

Future Directions

4.1. Few-Shot and Zero-Shot Learning

We train these models for Few-shot learning (classifying images with very few labelled examples) and zero-shot learning (using semantic information to classify unseen classes) [32]. These methods tackle what is perhaps most critical challenge in machine learning – data scarcity – while addressing data as it constantly emerges in dynamic fields such as personalized medicine, where new categories (e.g. rare diseases) are constantly being identified. Model-Agnostic Meta Learning (MAML) is an example of how meta learning algorithms, such as [33], are able to quickly adapt models to new tasks with very few data. Likewise, text descriptions of unseen categories can be used as such in zero shot learning methods such as employed in the CLIP model that are important for large scale, rapidly evolving tasks.

4.1.1 Progress in Few-Shot Learning

Meta learning is a type of learning such as Few-shot learning. That is a kind of process where we have a model that is able to learn in an autonomous way and improve in its performance by self-learning. It works like teaching a model to recognize things, or if you will, do tasks but it doesn't need to be overwhelmed with lots of examples, just few. The Few-shot learning seeks approaches to make the model learn quickly and efficiently from new and unseen data [33].

Machine leaning power Few-shot learning (FSL) has been proven a powerful approach to learn from little training examples. One of the offerings of this technique is that it is useful particularly in those cases where the time and cost of data collection is high. Here's how few-shot learning operates:

1. Learning from Minimal Data: Second, FSL generalizes from very few examples with the help of prior knowledge from related tasks. Meta learning strategy usually employed to learn how to optimize for new tasks with little data for performance.
2. Rapid Model Training: With Few-shot learning, they can train quickly and deploy quickly. It is

especially helpful in context of dynamic environments where we have multiple data requirements that change often.

4.2. Unsupervised and Self-Supervised Learning

In Unsupervised and self-supervised learning approaches, the learning is through unlabelled data directly using unsupervised and self-supervised learning techniques respectively. SimCLR and BYOL have promise [34]. These methods have also been particularly transformational in domains where labelled data is scarce or expensive: healthcare and astronomy. For instance, self-supervised learning has empowered important progress in medical imaging to automatically detect abnormalities from raw scans without annotated datasets. Unsupervised learning continues to be improved by techniques such as contrastive learning and masked image modelling which push the limitations of unsupervised learning techniques, allowing models to generalize better across diverse tasks.

4.2.1 Applications of Self-Supervised Learning

More recently, with the rise of self-supervised learning, the field of research in machine learning has witnessed a paradigm shift, and while promoting the use of unlabelled data, it has been proven to be quite a powerful paradigm [35]. It has wide ranging applications across computer vision, natural language processing, healthcare and robotics. Here are some notable applications:

1. **Computer Vision**
   - Image Classification and Object Detection: In SSL we pretrain models on large, unlabelled image datasets that are then fine-tuned for tasks like classification or detection (examples include SimCLR, MoCo).
   - Image Segmentation: It facilitates for pre-training of models for pixel level tasks, such as semantic segmentation.
   - Medical Imaging: It aids in tasks of anomaly detection, organ segmentation and disease diagnosis using very limited labelled medical data.
   - Face Recognition: Pre trained models help improving recognition accuracy and generalisation in the dataset we use.
2. **Natural Language Processing**
   - Language Models: SSL techniques are used by pre trained models such as BERT, GPT, RoBERTa to learn language representation from large text corpora.
   - Machine Translation: SSL pretraining helps to build translation systems on small amount of labelled data.
   - Text Summarization and Question Answering: All of the above pre-trained models improve downstream performance on text generation and comprehension tasks.
   - Sentiment Analysis: It comes into play in robust extracting feature for emotion and sentiment classification.
3. **Speech and Audio Processing**
   - Speech Recognition: Unlabelled audio data can be used to pre train SSL models like wav2vec and HuBERT that improve recognition systems.
   - Speaker Identification: It helps identify speakers in low resource scenarios.
   - Music Generation and Classification: It improves the ability to analyse and to generate music patterns.
4. **Robotics**
   - Reinforcement Learning: Learning representations for control tasks without extensive reward engineering is helped via SSL.
   - Autonomous Navigation: Provides insight for robots and autonomous vehicles into needed environment understanding and decision making.
   - Manipulation Tasks: Learns object manipulation with minimum labelled data.
5. **Recommendation Systems**
   - User Behaviour Modelling: Empirical results show that SSL pre-training improves the

prediction of user preferences in recommendation engines.

- Cold Start Problem: It generates initial recommendations when the data about the user is sparse.

4.3. Multimodal Learning

Due to the superior classification accuracy and expanded application domains that combining data from multiple modalities, e.g., text and images, can provide, this has become an important topic in research in recent years. For example, CLIP has shown the power of zero shot image classification through paired text image training [14]. More and more multimodal learning is applied in fields like medical diagnosis, bringing together the imaging data with patient records, to get the holistic view. Current models are now investigating multimodal fusion of video, audio and sensor data in the emerging realm of human activity recognition and autonomous systems that operate with ease within their respective environment. Some Key components of Multimodal Learning are listed below:

1. **Modalities:** It could be different types of data source or representation such as:
   - **Text:** For example, any type of natural language data (documents, social media posts).
   - **Images:** Data that can be visualized (e.g. photographs, medical scans).
   - **Audio:** Data that have sound (e.g., speech, music).
   - **Video:** Sequential visual data at hand (e.g., movies, surveillance footage).
   - **Other Sensors:** Devices such as LiDAR, accelerometers, EEGs.
2. **Fusion Techniques:** There are techniques for integrating information from multiple modalities including:
   - **Early Fusion:** Before the model processing, combining raw data or features extracted from all modalities.
   - **Late Fusion:** To combine outputs of unimodal models at decision making stage.
   - **Hybrid Fusion:** Performing both early and late fusion techniques to achieve better performance.
3. **Alignment:** Ensuring that two different modalities scale with each other to the same concepts or events: matching spoken words to video frames, for example.

4.3.1. Advances and Applications in Multimodal Learning

Diverse data types integration in multimodal learning would enhance the synthetic understanding of the environment. For example, imaging data can be combined with patient health records to make more accurate, and context aware, decisions in medical diagnostics. Currently advanced architectures explore dynamic fusion strategy which the model learns to weight the importance of each modality according to the task at hand. Emerging applications include:

- **Healthcare:** Using X-ray images in conjunction with the textual patient history in order to achieve better diagnostic accuracy.
- **Autonomous Systems:** Robust navigation by monitoring combined video feeds and sensor data along with GPS signals.
- **Human-Computer Interaction:** For real time decision making in assistive technologies, visual and auditory inputs.

4.4. Federated Learning

In Federated learning, a dataset is trained on distributed devices so that there is data privacy preservation [36]. On the meetup forums, there was also interest in this approach for sensitive applications such as healthcare, which cannot centralize patient data because of privacy. In supervised learning with medical imaging data from multiple hospitals, we used federated learning to train models without raw data sharing, improving diagnostic accuracy while preserving compliance with data protection regulations. One of the current research areas concentrates on how to address challenges such as handling heterogeneous data distributions across devices and decrease the communication overhead

during training. Along with improvements in security, scalable techniques such as secure aggregation and differential privacy are being integrated as well.

4.4.1. Challenges in Federated Learning

Distributed Machine Learning Paradigm Federated Learning is a way in which machine learning models can be trained on decentralized data without compromising privacy or security. In traditional machine learning, the data is collected, stored or processed in a single place or centralized [37]. However, sensitive information is easily collected, stored and processed, and this kind of data centralization creates privacy and security issues.

Federated learning is different from this, as machine learning models can be trained across decentralized data sources by providing the central server access to all other parties using the latter's local data to train distributed models but sharing the acquired parameters with the central server. By doing this, the server can summarize the models and then derives its global model after aggregating the models, while keeping the privacy and security of the data held.

Federated Learning (FL) is a distribution machine learning paradigm, where model training is done on decentralized devices and servers while keeping the data localized. Despite its potential, FL faces several challenges:

1. **Data Challenges**
   a. **Data Heterogeneity**
      - Non-IID Data: The data across devices is generally non independent and identically distributed (non IID), resulting in model divergence and poor performance.
      - Unbalanced Data: Depending on which device is involved, training imbalances can be far more severe.
      - Feature Skew: A solution that can degrade model performance due to various feature distributions across devices.

   b. **Data Privacy and Security**
      - Data Leakage: But even though I'm not sharing raw data, gradients and model updates can still leak sensitive information.
      - Differential Privacy Implementation: The problem of balancing the privacy guarantees with the model utility is complex.

2. **System Challenges**
   a. **Communication Overhead**
      - Bandwidth Constraints: When so large, large models can overwhelm networks by frequent communication involving the central server and devices.
      - Latency: Training can be delayed, due to slow or unstable network connections.

   b. **Device Resource Constraints**
      - Limited Computational Power: There are many devices (such as smartphones and lot of the IoT devices) with constrained processing power, memory and battery life.
      - Heterogeneous Devices: Depending on how many devices are used for training, and depending on varying hardware capabilities of the devices, then the training speed may be uneven and the resource utilization is unsynchronized.

   c. **Scalability**
      - Large-Scale Deployment: Logistical and computational challenges are raised by coordinating thousands or millions of devices.

3. **Algorithmic Challenges**
   a. **Model Aggregation**
      - Aggregation Bias: When we have non-IID data, federated averaging methods can potentially produce suboptimal global models.
      - Robustness: But it's also very difficult to ensure that the algorithm is robust against outliers or malicious updates.

b. Personalization

- Global vs. Local Models: If data is different across different devices, then a single global model can't work well.
- Personalized FL: It's complex developing algorithms that balance your generalization and personalization.

4.5. Quantum Machine Learning

In practical applications of Quantum computing for deep learning have just started [38], quantum computing can potentially accelerate deep learning acceleration. We expect quantum machine learning to have a revolutionary effect on computational efficiency, and specifically optimization, for large scale data processing. To date, as shown early experiments, quantum algorithms can speed up tasks like kernel-based classification and clustering. Indeed, hardware constraints are still limiting practical use cases, but there is ongoing research about integrating quantum circuits into conventional deep learning pipeline which might lead to hybrid quantum-classical models. Such advancements could drastically change the picture for computational biology, cryptography, and even financial modelling.

4.5.1. Theoretical Advances in Quantum Machine Learning

Quantum Machine Learning (QML) is a fast-growing merger of quantum computing and machine learning [39]. QML theoretical advances aim at using quantum mechanics to develop better machine learning algorithms and understanding the best and worst of quantum systems. Here are some significant theoretical advances in QML:

1. **Quantum Data Encoding**
   - Amplitude Encoding: Exponentially large feature spaces are represented efficiently as quantum states for classical data.
   - Quantum Feature Maps: Quantum maps classical data into high dimensional quantum Hilbert spaces to take advantage of quantum pattern recognition advantages.
   - Kernel Methods: The similarities in quantum feature spaces are computed via quantum kernels to improve classification and regression tasks.
2. **Quantum Speedups**
   - Exponential Speedup: Basic linear system solving, a fundamental ML operation, has known speedups using algorithms like the Harrow-Hassidim–Lloyd (HHL) algorithm.
   - Quadratic Speedup: Grover's search algorithms give quadratic speedups to optimization tasks in ML.
   - Sampling Speedup: Efficient sampling from distributions that are intractable for classical systems (e.g., quantum Boltzmann machine) is possible for quantum systems.
3. **Quantum Neural Networks (QNNs)**
   - Variational Quantum Circuits (VQCs): QNNs are modelled by parameterized quantum circuits. Quantum operations with classical optimization are combined.
   - Quantum Backpropagation: Training QNNs using gradient based methods theoretical frameworks.
   - Expressivity: We show that QNNs can express some functions more efficiently than classical neural networks.
4. **Quantum Support Vector Machines (QSVMs)**
   - Quantum Kernel Estimation: Other quantum algorithms like QSVMs compute kernels using quantum states, which are admitting of efficient handling of high dimensional data.
   - Advantage in High-Dimensional Spaces: Quantum feature spaces are exploited for improved classification in QSVMs.

Conclusion

Image classification has been transformed by Deep learning, reached unprecedented accuracy and opened doors to new ways of applying it. Unfortunately, data dependence, computational requirements, and ethical barriers remain. The topic

for future research is to develop efficient and interpretable models that also answer questions about fairness.

References

[1]. Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). "ImageNet Classification with Deep Convolutional Neural Networks." Advances in Neural Information Processing Systems. 2012, (1-10).

[2]. He, K., Zhang, X., Ren, S., & Sun, J. (2016). "Deep Residual Learning for Image Recognition." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016, (770–778).

[3]. Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, & Neil Houlsby. (2021). "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale." International Conference on Learning Representations.2021, (1-12).

[4]. Simonyan, K., & Zisserman, A. (2015). "Very Deep Convolutional Networks for Large-Scale Image Recognition." International Conference on Learning Representations.2021, (1-8).

[5]. Radford, A., Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, & Ilya Sutskever. (2021). "Learning Transferable Visual Models From Natural Language Supervision." International Conference on Machine Learning. 2021, (1–15).

[6]. Goodfellow, I., Jonathon Shlens, & Christian Szegedy. (2014). "Explaining and Harnessing Adversarial Examples." arXiv preprint arXiv:1412.6572. 2014, (1–10).

[7]. Kingma, D. P., & Welling, M. (2013). "Auto-Encoding Variational Bayes." arXiv preprint arXiv:1312.6114. 2013, (1–8).

[8]. Tan, M., & Le, Q. (2019) "Efficient Net: Rethinking Model Scaling for Convolutional Neural Networks." Proceedings of the International Conference on Machine Learning. 2019, (6105–6114).

[9]. Chen, T., Kornblith, S. Norouzi, M., & Hinton, G. (2020). "A Simple Framework for Contrastive Learning of Visual Representations." arXiv preprint arXiv:2002.05709. 2020, (1–12).

[10]. Esteva, A., Brett Kuprel , Roberto A Novoa , Justin Ko , Susan M Swetter , Helen M Blau , & Sebastian Thrun. (2017). "Dermatologist-level classification of skin cancer with deep neural networks." Nature. 2017, (115–118).

[11]. Zhu, X. X., Devis Tuia, Lichao Mou, Gui-Song Xia, Liangpei Zhang, & Feng Xu. (2017). "Deep Learning in Remote Sensing: A Comprehensive Review and List of Resources." IEEE Geoscience and Remote Sensing Magazine. 2017, (8–36).

[12]. Szegedy, C., Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, & Andrew Rabinovich. (2015). "Going Deeper with Convolutions." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015, (1–9).

[13]. Howard, A. G., Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, & Hartwig Adam. (2017). "Mobilenets: Efficient Convolutional Neural Networks for Mobile Vision Applications." arXiv preprint arXiv:1704.04861. 2017, (1–10).

[14]. OpenAI. (2021). "CLIP: Connecting Text and Images." arXiv preprint arXiv:2103.00020. 2021, (1–8).

[15]. Pan, S. J., & Yang, Q. (2010). "A Survey on Transfer Learning." IEEE Transactions on

Knowledge and Data Engineering. 2010, (1345–1359).

[16]. Yosinski, J., Jeff Clune, Yoshua Bengio, & Hod Lipson. (2014). "How Transferable Are Features in Deep Neural Networks?" Advances in Neural Information Processing Systems. 2014, (3320–3328).

[17]. Bello, I., Barret Zoph, Ashish Vaswani, Jonathon Shlens, & Quoc V. Le. (2021). "Attention Augmented Convolutional Networks." IEEE Transactions on Pattern Analysis and Machine Intelligence. 2021, (1–10).

[18]. Shorten, C., & Khoshgoftaar, T. M. (2019). "A survey on image data augmentation for deep learning." Journal of Big Data. 2019, (60–88).

[19]. Ioffe, S., & Szegedy, C. (2015). "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift." International Conference on Machine Learning. 2015, (448–456).

[20]. Yun, S., Dongyoon Han, Seong Joon Oh, Sanghyuk Chun, Junsuk Choe, & Youngjoon Yoo. (2019). "CutMix: Regularization Strategy to Train Strong Classifiers with Localizable Features." Proceedings of the IEEE International Conference on Computer Vision. 2019, (1–10).

[21]. Zhang, H., Moustapha Cisse, Yann N. Dauphin & David Lopez-Paz. (2018). "mixup: Beyond Empirical Risk Minimization." International Conference on Learning Representations. 2018, (1–12).

[22]. Opitz, D & Maclin, R. (1999). "Popular Ensemble Methods: An Empirical Study." Journal of Artificial Intelligence Research. 1999, (169–198).

[23]. Russakovsky, O, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein & Alexander C. Berg, Li Fei-Fei. (2015). "ImageNet Large Scale Visual Recognition Challenge." International Journal of Computer Vision. 2015, (211–252).

[24]. Zhu, X. J., & Goldberg, A. B. (2009). "Introduction to Semi-Supervised Learning." Synthesis Lectures on Artificial Intelligence and Machine Learning. 2009, (1–130).

[25]. Settles, B. (2012). "Active learning." Synthesis Lectures on Artificial Intelligence and Machine Learning. 2012, (1–114).

[26]. Han, S., Jeff Pool, John Tran, & William J. Dally. (2015). "Learning both Weights and Connections for Efficient Neural Networks." Advances in Neural Information Processing Systems. 2015, (1135–1143).

[27]. Selvaraju, R. R., Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, & Dhruv Batra. (2017). "Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization." Proceedings of the IEEE International Conference on Computer Vision. 2017, (618–626).

[28]. Miller, T. (2019). "Explanation in Artificial Intelligence: Insights from the Social Sciences." Artificial Intelligence. 2019, (1–38).

[29]. Goodfellow, I., Jonathon Shlens, & Christian Szegedy. (2015). "Explaining and Harnessing Adversarial Examples." International Conference on Learning Representations. 2015, (1–8).

[30]. Mehrabi, N., Fred Morstatter, Nripsuta Saxena, Kristina Lerman, & Aram Galstyan. (2021). "A Survey on Bias and Fairness in Machine Learning." ACM Computing Surveys. 2021, (1–38).

[31]. Kamiran, F., & Calders, T. (2009). "Classifying without Discriminating." International Conference on Computer, Control and Communication. 2009, (1–6).

[32]. Wang, Y., Quanming Yao, James Kwok, & Lionel M. Ni. (2020). "Generalizing from a Few Examples: A Survey on Few-Shot Learning." ACM Computing Surveys. 2020, (1–34).

[33]. Snell, J., Kevin Swersky, & Richard S. Zemel. (2017). "Prototypical Networks for Few-shot Learning." Advances in Neural Information Processing Systems. 2017, (4080–4090).

[34]. Grill, J. B., Florian Strub, Florent Altché, Corentin Tallec, Pierre H. Richemond, Elena Buchatskaya, Carl Doersch, Bernardo Avila Pires, Zhaohan Daniel Guo, Mohammad Gheshlaghi Azar, Bilal Piot, Koray Kavukcuoglu, Rémi Munos, & Michal Valko. (2020). "Bootstrap Your Own Latent: A New Approach to Self-Supervised Learning." Advances in Neural Information Processing Systems. 2020, (1–12).

[35]. He, K., Haoqi Fan, Yuxin Wu, Saining Xie, & Ross Girshick. (2020). "Momentum Contrast for Unsupervised Visual Representation Learning." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020, (9729–9738).

[36]. Kairouz, P., H. Brendan McMahan, Brendan Avent, Aurélien Bellet, Mehdi Bennis, Arjun Nitin Bhagoji, Kallista Bonawitz, Zachary Charles, Graham Cormode, Rachel Cummings, Rafael G.L. D'Oliveira, Hubert Eichner, Salim El Rouayheb, David Evans, Josh Gardner, Zachary Garrett, Adrià Gascón, Badih Ghazi, Phillip B. Gibbons, Marco Gruteser, Zaid Harchaoui, Chaoyang He, Lie He, Zhouyuan Huo, Ben Hutchinson, Justin Hsu, Martin Jaggi, Tara Javidi, Gauri Joshi, Mikhail Khodak, Jakub Konečný, Aleksandra Korolova, Farinaz Koushanfar, Sanmi Koyejo, Tancrède Lepoint, Yang Liu, Prateek Mittal, Mehryar Mohri, Richard Nock, Ayfer Özgür, Rasmus Pagh, Mariana Raykova, Hang Qi, Daniel Ramage, Ramesh Raskar, Dawn Song, Weikang Song, Sebastian U. Stich, Ziteng Sun, Ananda Theertha Suresh, Florian Tramèr, Praneeth Vepakomma, Jianyu Wang, Li Xiong, Zheng Xu, Qiang Yang, Felix X. Yu, Han Yu, & Sen Zhao. (2019). "Advances and Open Problems in Federated Learning." arXiv preprint arXiv:1912.04977. 2019, (1–21).

[37]. Li, T., Anit Kumar Sahu, Manzil Zaheer, Maziar Sanjabi, Ameet Talwalkar, & Virginia Smith. (2020). "Federated Optimization in Heterogeneous Networks." Proceedings of Machine Learning and Systems. 2020, (429–450).

[38]. Biamonte, J., Peter Wittek, Nicola Pancotti, & Patrick Rebentros (2017). "Quantum Machine Learning." Nature. 2017, (195–202).

[39]. Schuld, M & Petruccione, F. (2018). "Supervised Learning with Quantum Computers." Springer. 2018, (1–300).

[40]. Jagdish Jangid , " Efficient Training Data Caching for Deep Learning in Edge Computing Networks" International Journal of Scientific Research in Computer Science, Engineering and Information Technology(IJSRCSEIT), ISSN : 2456-3307, Volume 6, Issue 5, pp.337-362, September-October-2020. Available at doi : https://doi.org/10.32628/CSEIT20631113