A Review of Convolutional Neural Networks and Gabor Filters in Object Recognition

Mehang Rai* and Pablo Rivas†, *Senior, IEEE*
School of Engineering and Computer Science
Department of Computer Science, Baylor University
*Email: Mehang_Rai1@Baylor.edu, †Email: Pablo_Rivas@Baylor.edu

***Abstract*—Convolutional neural networks (CNNs) have become a classic approach to solving challenging computer vision problems. Much of its success is due to its ability to discover optimal filters that capture non-trivial spatial relationships in data. Other vital components include advances in optimization, regularization, and overfitting prevention strategies. However, recently, researchers have observed closely the connections between what CNNs learn in the layers that capture low-level features and filter-banks such as Gabor filters. Gabor filters have been used in computer vision tasks long before CNNs were popularized with good performance. This paper presents a review of the literature concerning approaches that involve both Gabor filters and CNNs. We pay close attention to successes and opportunities for future research in the intersection of these two computer vision tools.**

***Index Terms*—convolutional neural networks, deep learning, Gabor filters, object recognition, image processing**

I. Introduction

It is interesting how human beings are able to recognize an object. First, our eye perceives the image of an object. Then, the brain extracts the details from the image and aggregates a concise result from those details. This intricate process of object recognition simply goes unnoticed as our eyes and brain coordinate in such a subtle way. Because the researchers are fascinated by this, object recognition has been a highly contemplated topic in the field of computer vision for decades. Although researchers have already explored its vast range of topics, there is still room for improvement. The progress may be gradual, but with the advent of new and improved algorithms and technologies, the progress seems inevitable.

Talking about image processing, it is impossible to skip over Gabor filters. In the field of image processing, it has garnered well reputation regarding its use. The very concept our eye depends on to detect an object is recognizing the shape and structure of the object based on the texture segmentation, and this segmentation is the main essence of Gabor filters. Various research has shown its efficient usage in different scenarios [1], [2], [3], [4], [5], [6], [7].

But in the last decade, Convolutional Neural Networks (CNNs) have been favored extensively over Gabor filters. Although the usage of CNN dates back to the late 1990s [8], it has again burst into the scene since the early 2010s, and this can be hugely credit to the initial effort of Krizhevsky, Stuskever, and Hinton [9]. Since their work, many variations of the model has been approached in many different sectors [10], [11], [12], [13], [14], [15].

Because of the technical limitation, in the past, CNN was not considered as the de-facto model. But, now with the availability of surging computation power in Central Processing Unit (CPU), Graphics Processing Unit (GPU), and even in cloud computing, the limitation has been alleviated, and this has certainly favored CNN. And why not? It does not require one to have the expertise of parameters of Gabor filters. It learns through its way through all the learning space, optimizing itself to give the desired result. But, granted the power of CNN in self-optimization, we believe the essence of the Gabor filter not to be forgotten. There have been some research done, forming the bond between these two powerful models [16], [17], [18] and it has resulted in some pretty interesting result to look upon. These results have made us ponder upon the different possibilities that it could lead to. With the presented summary of all the research over the past decade, we certainly hope to motivate the reader to pursue research in the intersecting fields of CNNs and Gabor filters and its symbiotic relationship.

The rest of the paper is organized as follows: Section II discusses Gabor filters and their existing use across different fields. Then we will be talking about CNNs, its basic concept, different variations of its model, and its effective use in a variety of sectors in Section III. Then, Section IV addresses how Gabor filters and CNNs have been used to augment the productivity of learning process. Finally, Section V presents a brief discussion and concluding remarks are drawn in Section VI.

II. Gabor Filter

A Gabor filter, derived from Gabor elementary functions (GEF), is a linear filter used for a multitude of image processing applications for texture analysis, edge detection, feature extraction, etc. As a band-pass filter, the Gabor filter enables the extraction of patterns at the specified certain frequency and orientation of the signal. Therefore this resulting property of transforming texture differences into detectable filter-output discontinuities at texture boundary has established itself to mimic the functionality of the visual cortex [1], [2], [3]. While the concept of Gabor elementary function was initially presented by Hungarian-British physicist Dennis Gabor [19], it was later extended to 2-D filters by Daugman [20].

In basic terms, a Gabor elementary function (GEF) can be thought of as a Gaussian being modulated by a complex

978-1-7281-7624-6/20/$31.00 ©2020 IEEE
DOI 10.1109/CSCI51800.2020.00289

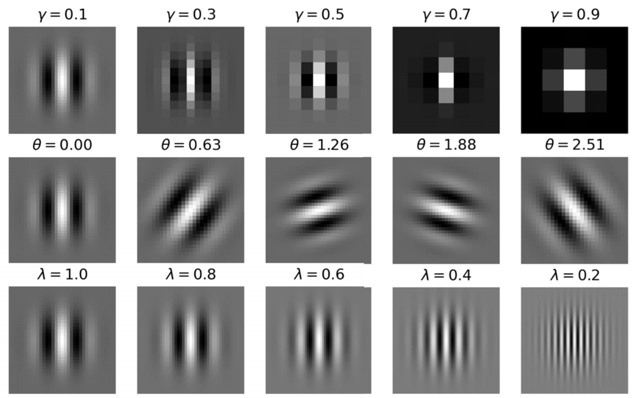

Fig. 1. Different Gabor filters with different values for λ, θ and γ. Different parameters will change filter properties.

sinusoid, where the Cosine and Sine waves generate the real and imaginary component. GEF can be formulated as:

$$g(x, y) = \exp\left(-\frac{x'^2 + \gamma^2 y'^2}{2\sigma^2}\right) \exp\left(i\left(2\pi\frac{x'}{\lambda} + \psi\right)\right) \quad (1)$$

where $(x', y') = (x\cos\theta + y\sin\theta, -x\sin\theta + y\cos\theta)$ represents rotated spatial-domain rectilinear coordinates; λ represents the wavelength of the sinusoidal factor, θ represents the orientation of the normal to the parallel stripes of a Gabor function, ψ is the offset, σ_x and σ_y characterizes the spatial extent and bandwidth of the filter. Most of the time a symmetric filter ($\sigma_x = \sigma_y$) will suffice for texture segmentation, but when the texture contains texels not arranged in square lattice, asymmetric filters ($\sigma_x \neq \sigma_y$) could be useful [1]. This asymmetric nature is given by $\gamma \neq 1$, where γ is the spatial aspect ratio:

$$\gamma = \frac{\sigma_x}{\sigma_y}. \quad (2)$$

Since a single Gabor filter can only be responsible for a certain feature, a multitude of Gabor filters is necessary to yield meaningful features. As shown by Jain *et al.* [3] a multitude of features computed over different spatial orientations and frequencies was necessary to yield a successful segmentation of an image with complex background. Because of its reputation in texture segmentation, Gabor features have also been popular in the automated defect detection of textured materials [21], [22], [7]. But, since the Gabor filter only gives out texture features, an algorithm that yields a meaningful result from those features is necessary. Kumar and Sherly [21], used a multi-channel filtering scheme, while Jing *et al.* [22] used Kernel Principal Component Analysis upon the extracted features along with the OTSU threshold method to give a high defect detection rate. Correspondingly, Li *et al.* [7] used Pulse Coupled Neural Network (PCNN) giving a detection accuracy of around 98.6%

Over the years, the Gabor filter has seen its use in a variety of applications. Li *et al.* [7], proposed a method for road detection using Gabor filter. They effectively demonstrated the robustness of Gabor filters by detecting roads in various lighting conditions (night, entering tunnel, and shadowing). Their proposed methods consisted of two steps: locating vanishing-point based on soft voting scheme upon dominant texture orientations, and then detecting the road lane ahead of the vehicle via edge detection method, while effectively constraining the search of lane mark using the vanishing point.

El-Sayed *et al.* [23] proposed an authentication mechanism based on the identification of retinal features. They efficiently used the Gabor filter to segment the retinal blood-vessel, and then ran SVM was upon the resulting feature pattern for feature matching. They claim this method to be stable regarding multiple and rotary shifts of digital retina images, and their test result corroborates their claim as they were able to achieve an accuracy of around 96.9%.

In their research, Gornale *et al.* [24] showed an interesting way of identifying gender-based on features gathered from Discrete Wavelet Transform (DWT) and Gabor based feature. When most of the research was focusing on facial features, this was a pretty interesting method as it was able to achieve 97% accuracy.

In 2016, Rizvi *et al.* [25] demonstrated the use of Gabor features for object detection. Aided with Gabor filters, the feedforward Neural Network model was able to an accuracy of 50.71%, which was comparable to that of CNN (52.15%). This was interesting as it was able to achieve such accuracy in less amount of training time.

Avinash *et al.* [26], argued about the failure of previously employed methods in a real-time application for detection of lung cancer in early stages. They proceed on to propose the usage of Gabor filter along with Marker driven watershed segmentation technique on Computed Tomography (CT) images to overcome the hurdle.

Continuing on Gabor filters, Daamouche *et al.* [27] proposed an unsupervised method of application of Gabor filters and morphological operators for building detection on remotely sensed images.

Over the years, the Gabor filter has seen its heavy usage in the extraction of facial features. In 2016, Hemalatha and Sumathi [28] proposed the Median and Gabor filters along with Histogram Equalization as a combined preprocessing method, for yielding a better-enhanced image. They argue that their technique will lead to a color-normalized, noise-reduced, edge-enhanced, and contrast illuminated image.

In the same fashion, Lefkovits *et al.* proposed the use of Gabor filters to aid detection of the eye and its openness [29]. Their methodology primarily consisted of using the Gabor filter to detect the eye which was aided with Viola-Jones face detection [30] to speed up the process and a self-created face classifier based on Haar features to lower false positive of detection rate.

Around the same time, Pumlumchiak and Vittayakorn [31] presented a novel framework for facial expression recognition. Their method primarily consisted of extraction of Gabor filter responses as facial features, mapped upon feature subspace using the joint framework of Principal Component Analysis

(PCA), Principle Components (PCs) removal, and Linear Discriminant Analysis (LDA). Their experimental result shows to outperform existing baselines, and thus substantiating that the weighted neighbor to be a good approach at classifying facial expressions to 4 different classes: anger, surprise, happiness, and neutral.

However, Mahmood *et al.* [32] went with a different approach in tackling the facial expression recognition task. Their method comprised of a combined Radon transform and Gabor transform for facial feature extraction, fed towards a fused-classifier approach in the form of Neural Network over Self-Organized Maps (SOM). On classification over 6 different expressions - surprise, anger, sadness, disgust, happiness, and fear, an accuracy of 84.87% was obtained over two public datasets, on average.

Rather than using Standard Gabor Filter Ensemble (SGFE) of varying scale and orientation, Low *et al.* [33] proposed a Condensed Gabor Filter Ensemble (CGFE) in which have the diversified traits of multiple SGFE are condensed into a single one. Their method of self-cross convolving the pre-selected Gabor filters exhibit to outperform the state of the art face descriptors Linear Binary Pattern (LBP) variants: Discriminant Face Descriptor (DFD) [34] and Compact Binary Face Descriptor (CBFD) [35].

Nava *et al.* [36], in 2012, proposed a new filtering scheme, Log-Gabor, designed to eliminate the non-uniform coverage in Fourier domain produced by Gabor filter, and thus strongly correlating with Human Visual System (HVS). In 2017, Nunes *et al.* [5], expanded on this filtering scheme and proposed a local descriptor called multi-spectral feature descriptor (MFD), designed specifically to work with images acquired over different frequencies across the electromagnetic spectrum. Upon evaluation, it was found to be computationally efficient while maintaining the same precision and recall as the extant state-of-the-art algorithms.

The feature point matching method presented by Liu *et al.* [37] for infrared and visible image matching also effectively utilizes Log-Gabor for generating the descriptors. This method based upon the Log-Gabor filters and Distinct Wavelength Phase Congruency(DWPC) effectively helps in matching non-linear images with different physical wavelength, and the experiment results corroborate it, as this method outperformed traditional approaches: edge-oriented histogram descriptor (EHD), phase congruency edge-oriented histogram descriptor (PCEHD), and log-Gabor histogram descriptor (LGHD), in infrared and visible images by 50%.

In the case of image segmentation, the Gabor filter is always the one to look up to, and Premana *et al.* [6] demonstrated this by using just simple, yet powerful K-Means clustering algorithm to segment the object from its complex background with the aid of Gabor filter responses. Fan *et al.* [38] proposed a novel woven fabric recognition method based on a similar concept. They proposed the utilization of Gabor filter to determine the orientation of texture at yarn crossing points segmented with K-means clustering and gradient accumulation. This segmentation capability of the Gabor filter is further demonstrated by Gargi Srivastava and Rajeev Srivastava [39] as they propose a novel method for salient object detection. Combined with the foreground saliency map formed from backgroundness score via minimum directed backgroundness and segmented images obtained from Gabor filters, this method utilizes an objectness criterion to choose the segment containing the salient object. Although failing in some conditions, this method effectively outperforms state-of-the-art algorithms (evaluated by PR-curve, F-Measure curve, and Mean Absolute Error upon 8 different public datasets).

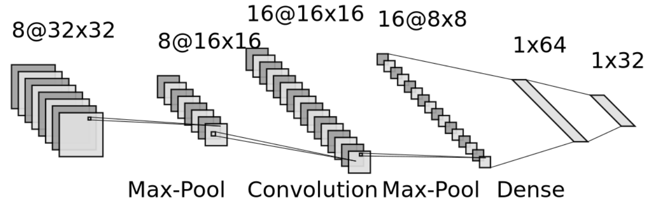

Fig. 2. An illustration of architecture of CNN with convolutional layers, pooling layers and dense layers.

Recently, in 2019, Khaleefah *et al.* [40] proposed an interesting method to combat the deformations in paper images formed by extant scanners. Their novel Automated Paper Fingerprinting (APF) utilized the combined effort of Gabor filters and Uniform Local Binary Patterns (ULBP) for extracting both local and global information for better texture classification. Their evaluation effectively highlights the need for Gabor filter as the combined approach was able to outperform the standalone ULBP system by 30.68%.

III. Convolutional Neural Network

In the field of image processing, many consider Convolutional Neural Network (CNN) to be state-of-the-art. CNN represents a family of statistically learning models which is primarily based upon the convolution operation of images with filters leading to feature-mapping layers. In a similar fashion to any other Neural Network (NN) models, it is biologically inspired by visual neuroscience theory. Hubel and Wiesel [41], found out that in a cat's visual cortex there are simple cells and complex cells present which fire in response to certain properties of visual sensory inputs. While the simple cell showed a response to simple, low-level spatial features like the orientation of edges, complex cells exhibited more spatial invariance. And, the architecture of CNN is similar to that - a hierarchical multi-layer network where receptive layers are designed to capture some specific peculiarity of the image while the following layers build upon that to create more abstract features.

A general CNN consists of some combinations of convolution layers, pooling layer, activation layer, and dense (fully-connected) layer, but modification can be seen according to the application in-hand like the addition of dropout layer, normalization layer, etc. for issues like overfitting, uniformity, etc. CNN is trained usually via backpropagation [42] in which the weight is updated using variation gradient descent

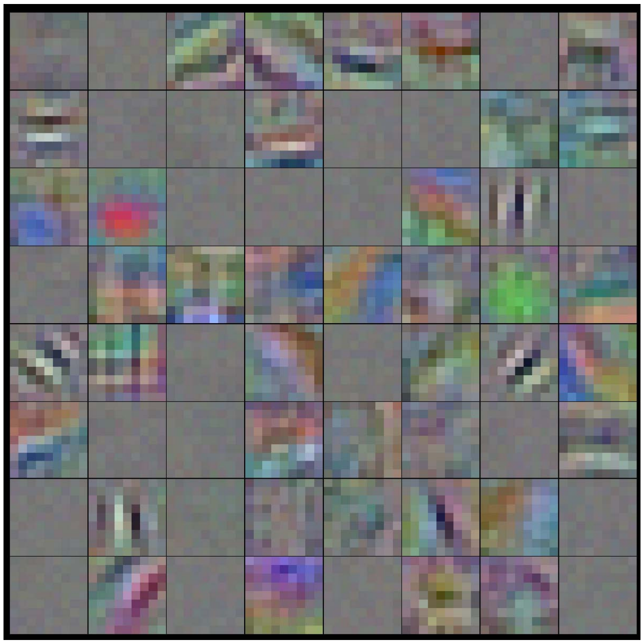

Fig. 3. Visual representation of 64 convolutional kernels of size 9×9×3 learned by the first convolutional layer on the 32×32×3 input images.

like Stochastic Gradient Descent (SGD), Mini-batch Gradient Descent, etc. [43].

CNN is not new in the field. It has seen its fair share back in the 1990s too. Cun *et al.* [42] effectively utilized the model for recognizing handwritten zip codes to constrain the error rate to 1%, while Lawrence *et al.* [8] and Rivas and Chacon [44] effectively demonstrated the capability of CNN in face recognition by outperforming Karhunen-Loeve (KL) transform and Multi-Layer Perceptron (MLP). But it was the work of Krizhevsky *et al.* [9] that brought CNN back into the limelight. AlexNet made a significant stride in the field of image recognition as it effectively demonstrated that a deeper model is much better than a wider model. With a margin of 10.9% compared to the second-placed model, it won the ILSVRC-2012 competition. Following the success of AlexNet various deeper models like Residual Networks (ResNets) [45], GoogleNet [46], VGGNet [47], etc. has came into existence, and also various research like object recognition [48], [49], [50], [51], 3D object detection [52], [53], pedestrian detection [13], learning scene gist [54], etc. has been conducted.

CNN generally outperforms other supervised learning algorithms when it comes to image processing. In [48], researchers confirmed CNN outperforms Support Vector Machine (SVM) while automating the task of image analysis coming from the satellite. Szarvas *et al.* [13] showed that CNN is able to reduce the False Positive Rate (FPR) to less than $\frac{1}{5}$ of SVM when trained on pedestrian images with complex background and subject, and they mainly attributed this to the optimization of feature representation by CNN. Likewise, regarding the multi-class object recognition problem, Hayat *et al.* [55] showed that a 5-layered CNN was able to achieve 90.12% accuracy, which completely outperformed different classical bag-of-words (BOW) approaches. Similarly, Zulkeflie *et al.* [51] evaluates AlexNet, basic CNN, and Bag of Features (BoF) with Speeded-Up Robust Feature (SURF) and SVM classifier, and found out that AlexNet and basic CNN model outperforms BoF model. Looking at all this research, it can be seen that most of the time, the general CNN model suffices for the task at hand, sometimes some tweaks may be necessary, and it may be concerning the complexity of task, accuracy, memory requirements, etc.

In 2014, Kawano and Yanai [12] integrated conventional hand-crafted image features, namely, Fisher Vectors with Histogram of Oriented Gradients (HOG) and Color patches, with the convolutional features, boosting the accuracy to 72.6% in a 100-class food dataset. It completely outperformed the existing best accuracy rate - which was at 59.6%.

Biologically inspired, Wu *et al.* [54] proposed integration of scene's gist for object recognition improvement, similar to how humans foveate on an object and incorporate periphery information to aid object recognition. Coined as GistNet, their model consisted of two CNN models - a fovea sub-network for object recognition and a periphery sub-network for contextual modulation. With VGG-16 as a baseline, their approach improved the accuracy by 50% for certain object recognition while increasing the size by only 5%.

Kumar and Sherly [49] fine-tuned the last two layers of a pre-trained VGG-16 CNN model and trained on their augmented data to avoid overfitting due to lack of training data. This approach led them to an accuracy of 81.6%, which is good considering the scarcity of training data.

Talking about overfitting and data deficiency, as a supervised learning approach, CNN needs a large amount of data in order to boost its performance and generalization. While transfer learning could be done to deviate from the need for an expensive labeling process, Dosovitskiy *et al.* [56] proposed a discriminative unsupervised feature learning approach. Training the network to discriminate between surrogate classes, created by applying a variety of transformations to a randomly sampled seed image path, led it to outperform extant state-of-the-art unsupervised methods. With regard to the context and argument, this novel approach is certainly to look up to.

Over the years, CNN has seen its fair share of use as a feature extractor too [57], [58], [52], [59], [60]. Upon evaluation of AlexNet and VGGNet, researchers in [58] showcased that not only the final fully-connected layers but the intermediate layers too, can act as a source of features to enhance recognition performance. Similarly, Chen *et al.* [57] also ascertained that different layers in CNN could engender features suitable for the detection of different aspects of place recognition task.

In 2015, Wang *et al.* [52], proposed the use of CNN along with SVM, where CNN acts as the feature extractor and SVM as the classifier, for 3D object recognition. In their proposed approach, they first converted depth modality into 3 channels, and then fine-tuned two pre-trained Caffe models [61], in order to extract representative sparse features from color (RGB) and depth (RGB-D) images, finally to be used by SVM classifier. On experimentation, this approach yielded 91.35% accuracy, much better than the state-of-the-art and also CNN model trained solely on RGB images from the RGB-D object dataset.

Similarly, Schwarz *et al.* [53] proposed the use of a pre-trained CNN model as a feature extractor, in conjunction with SVM classifier for RGB-D object recognition its pose estimation. Their approach incorporated depth features by rendering objects from canonical views and coding metric distance from the object center with the color scheme, making it well suitable for CNN to extract meaningful features.

Continuing on 3D object recognition, Gao *et al.* [62] proposed pairwise Multi-View CNN (coined as PMV-CNN), designed to explicitly deal with lack of training samples while also maintaining latent complementary information from different views explored via view pooling. Their novel approach used a pair of CNN in order to jointly learn the visual features from multiple views and optimize towards object recognition.

Since Spiking Neural Network (SNN) based architecture is energy efficient when used in conjunction with spike-based neuromorphic hardware,Cao *et al.* [59] proposed a novel approach for converting CNN into an SNN in order to map into spike-based hardware. When training the tailored CNN, this approach gets exposed to the learning capability of CNN, and while transferring the learned weight of the tailored CNN back to SNN, it becomes energy efficient and compatible with spike-based neuromorphic hardware. Regarding real-time object recognition, they found this approach to be energy efficient than Field Programmable Gate Array (FPGA)-based implementation of CNN by two orders of magnitude.

As shown by all these aforementioned research, while CNN has established as the state of the art for object recognition, it can be expanded to recognition in real-time too. Upon implementation of CNN on FPGA, Ahn [63] was able to achieve 170,000 classifications per second and scale-invariant object recognition from a 720×480 video stream at a speed of 60 fps. Similarly, Radovic *et al.* [64] proposed the use of YOLO - a CNN based open-source object detection and classification platform - for classification of the object on real-time video feed obtained form Unmanned Aerial Vehicles (UAV).

In [14], Maturana and Scherer proposed a 3D CNN architecture, coined VoxNet, that integrated a volumetric occupancy grid representation with 3D CNN for real-time object detection. This representation enabled full utilization of information coming from range sensors, ultimately boosting performance to labeling hundreds of instances per second. Inspired by [14], Garcia-Garcia *et al.* [65] proposed the use of density occupancy grids as the inner representation for input data in a model coined PointNet. When integrated with the 3D CNN model, this approach significantly boosted the performance. Expanding upon [14], Zhi *et al.* [15] proposed LightNet - a lightweight volumetric 3D CNN. Their compact model was computationally efficient that VoxNet, while a combination of different kinds of auxiliary learning tasks made it less vulnerable to overfitting.

Likewise, Huang and You [66] introduced a 3D point cloud labeling scheme based on 3D CNN. Representation based on only voxelized data made it straightforward. While complications like exceeding memory usage, biased classification, etc. could exist, they did present solutions for handling such data.

In [60], Fang *et al.* devised a novel approach Improved Faster Regions with CNN Features (IFaster R-CNN) to address the generalization issue while detecting objects on construction sites in real-time. Their approach was also based upon the use of CNN as base feature extractor from images, which then with the use of Region Proposal Network (RPN) to concurrently predict object bounds and objectness scores at a particular position, fed the extracted regional proposals were fed into Fast R-CNN module for detection. With detection speed at real-time at 0.101 s per image and accuracy of about 91% and 95% for worker and excavator respectively, they completely outperformed extant state-of-the-art by an average of 50%.

Du *et al.* [67] experimented with a six-degree-of-freedom (6-DOF) robot arm with a gripper, their proposed method successfully yielded an accuracy of 98.44% for the stereo vision-based object recognition and manipulation. Their hybrid algorithm comprised of an adaptive network-based fuzzy inference system (ANFIS) for the eye-to-hand calibration and R-CNN for object detection.

While accuracy has been the most important aspect researchers looked up to, there has been considerable research done to boost the speed of recognition too [11], [10], [68]. The authors of [11] were able to considerably drop the error rates in fewer epochs when trained using their fast, fully parameterizable GPU based CNN.

Similarly, with regard to speed, Anwar *et al.* [10] proposed fixed-point optimization for reducing the number of parameters. They effectively quantized layers of pre-trained high precision networks using L2 error minimization based on layerwise sensitivity on word-length reduction. Their approach not only significantly reduced memory usage but also generalized the model.

As 3D object detection is computationally demanding, Xu *et al.* [68] proposed Volumetric Accelerator (VOLA) for the memory-efficient representation of the 3D volumetric object. With a reduction in memory usage, they purport their representation model to be better in terms of speed, and their experimental result advocates it as their VOLA-based CNN performed 1.5 times faster than the original LeNet.

IV. Gabor and CNN

Judging from all these research it can be clearly seen that both Gabor filter and CNN can act as an excellent feature extractor. However as seen in previous research [12], [54], CNN can skip over some of the valid specific information and hence, can immensely benefit when complemented with other manual features. Since Gabor has been defined to extract all sorts of features [31], [29], [5] in a different domain, this makes it a well-suited candidate to complement CNN, and in fact, a lot of research have shown so [69], [18], [4], [70].

In conjunction with Gabor filter features, Yao *et al.* [69] found that CNN yielded a 1.26% boost in accuracy when employed for object recognition in the natural scene. With an accuracy of 81.53%, it outperformed standalone CNN marginally and significantly outperformed the Bag-of-Words model with Scale Invariant Feature Transform (SIFT). In the same manner, Hosseini *et al.* [4] utilized the Gabor filter responses to boosting the accuracy of CNN for classification based on age and gender. Zadeh *et al.* [71], also noted the boost in speed and accuracy when Gabor filter features were incorporated with CNN for fast facial emotion recognition.

The visualization of Gabor filters and first convolutional layers of CNN shows that they are quite alike, and it was confirmed by Krizhevsky *et al.* [9] that when trained on real images deep CNN the first convolutional layers to be similar to Gabor filters. Motivated by this fact, Alekseev and Bobe [72] modified the architecture where the first layer of CNN was constrained to fit the Gabor function. Upon experimentation with different datasets, it was found to yield the same or even better accuracy with significant improvement in convergence.

Inspired by traditional local Gabor binary patterns, Jiang and Su [70] proposed Gabor Binary Layer (GBL) as an alternative for the first layer of the CNN model. GBL - composed of a module of predefined Gabor filters with different shape and orientation and a module of fixed randomly generated binary filters - when experimented with different CNN models gave a better performance than the state-of-the-art CNNs.

In a similar fashion, Luan *et al.* [16] extended the concept to multiple layers CNN. Coined as Gabor Convolutional Networks (GCN), their network comprised of predefined Gabor filters of different scale and orientation in multiple layers. Proposed to enhance the robustness of the model against image transitions, scale changes, and rotations, their model significantly enhanced performance over the baseline model while simultaneously reducing the training complexity too.

Built upon GCN, Liu *et al.* [17] proposed a new learning model, Hybrid Gabor Convolutional Network (HGCN). While [16] focused on accuracy, [17] went for memory efficiency. With hybrid binarized input and Gabor Binarized Filters (GBFs) in an end-to-end framework, HGCN was able to reduce memory usage by a factor of 32 while maintaining accuracy due to usage of GCN.

Molaei *et al.* [18] also initialized the first layer of CNN with predefined Gabor filters for effective Left Ventricle segmentation. Due to the robust nature of the Gabor filter, the model increased the performance in terms of specificity and sensitivity. In 2020, Molaei and Shiri Ahmad Abadi [73], expanded the model to maintain the structure of the Gabor filter during the training process. When compared with different initialization methods, it significantly outperformed all, even when dealing with noisy data and a lesser amount of training data.

V. Discussion

Gabor filters are an established methodology to capture image properties using frequency-domain theory that has direct feature extraction properties in the spatial domain. Clearly, over the years Gabor filters were the method of choice for general purpose feature extraction on image-based object recognition tasks. However, when CNNs were designed to lear any sort of filter, which includes the set of all possible Gabor filters, then researchers quickly abandoned its use in favor of optimal filter design through CNNs.

Recenlty, with the stability of CNNs researchers are now looking closely at what CNNs are learning and have desired to investigate how CNNs work and why. The research presented here, points out the direct possible relationship between CNNs and Gabor filters and calls researcher to pay attention to this new and exciting area. The work presented here summarizes research conducted using Gabor filters in conjunction with CNNs to improve accuracy.

The research includes both CNNs and Gabor filters have shown promising results when the filter parameters are guided through back-propagation; however, this literally removes a CNNs ability to arbitrarily destroy a filter if that is what is needed to find a local minima. Therefore, it must be worth noting that the ability for a CNN to fully alter a filter's design must be kept as this has bee proven to be one of its mayor benefits.

However, we claim that to the best of our knowledge there are no algorithms that initialize CNNs with Gabor filters and let them be freely modified with back-propagation, rendering Gabor filters as an initial tool that may or may not need to be forced into the network. Although, deeper convolutional layer can develop the ability to ignore and alter the information given by Gabor filters, these modification have exponential complexity in terms of the amount of updates that are required.

VI. Conclusion

CNNs have become the preferred method for computer vision problems aiming for classification, clustering, and other image analysis tasks. However, general-purpose object recognition tasks often seem to yield low-level features closely related to Gabor filters.

The literature examined shows that when Gabor filters have been used in computer vision tasks, performance is often superior compared to other approaches that do not use such an approach. The computational expense of calculating Gabor filters is negligible, having constant-time complexity, $O(1)$, implying easy acquisition and deployment.

For these reasons, researchers have recently combined the Gabor filter theory and CNNs. The results have been positive in specific computer vision tasks, allowing gradient descent techniques to update specific subsets of the set of parameters of the Gabor filters, $\{\gamma, \lambda, \theta\}$, in order to use the pseudo-optimal collection of filters.

However, existing approaches have a number of issues that have not been explored and that we will tackle in further research. First, restricting Gabor filters as the only thing that a CNN can use might be severely limiting the potential of a CNN to alter the structure, even so slightly, of a Gabor filter in order to maximize performance, or even completely destroy the spatial shape of an existing, under-performing, filter. Second, current studies combining Gabor filters and CNNs have not shown a conclusive relationship between the use of Gabor filters and the convergence of a CNN, which is crucial to understanding the added computational cost of using Gabor filters as opposed to using randomly generated uniform white noise, which is the traditional approach. Third, while the evidence that CNNs and Gabor filters together are successful in very specific computer vision tasks, there is no sufficient evidence that Gabor filters can provide a significant advantage in general object recognition tasks. We intend to tackle these and other consequential issues in our on-going research.

Acknowledgment

The authors would like to thank the support of the Department of Computer Science at Baylor University during this research study.

References

[1] D. Dunn, W. E. Higgins, and J. Wakeley, "Texture segmentation using 2-d gabor elementary functions," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 16, no. 2, pp. 130–149, 1994.

[2] D. Dunn and W. E. Higgins, "Optimal gabor filters for texture segmentation," *IEEE Transactions on Image Processing*, vol. 4, no. 7, pp. 947–964, 1995.

[3] A. K. Jain, N. K. Ratha, and S. Lakshmanan, "Object detection using gabor filters," *Pattern Recognition*, vol. 30, no. 2, pp. 295 – 309, 1997. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0031320396000684

[4] S. Hosseini, S. H. Lee, H. J. Kwon, H. I. Koo, and N. I. Cho, "Age and gender classification using wide convolutional neural network and gabor filter," in *2018 International Workshop on Advanced Image Technology (IWAIT)*, 2018, pp. 1–3.

[5] C. F. G. Nunes and F. L. C. Pádua, "A local feature descriptor based on log-gabor filters for keypoint matching in multispectral images," *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 10, pp. 1850–1854, 2017.

[6] A. Premana, A. P. Wijaya, and M. A. Soeleman, "Image segmentation using gabor filter and k-means clustering method," in *2017 International Seminar on Application for Technology of Information and Communication (iSemantic)*, 2017, pp. 95–99.

[7] Z. Li, H. Ma, and Z. Liu, "Road lane detection with gabor filters," in *2016 International Conference on Information System and Artificial Intelligence (ISAI)*, 2016, pp. 436–440.

[8] S. Lawrence, C. L. Giles, Ah Chung Tsoi, and A. D. Back, "Face recognition: a convolutional neural-network approach," *IEEE Transactions on Neural Networks*, vol. 8, no. 1, pp. 98–113, 1997.

[9] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems 25*, F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, Eds. Curran Associates, Inc., 2012, pp. 1097–1105. [Online]. Available: http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf

[10] S. Anwar, K. Hwang, and W. Sung, "Fixed point optimization of deep convolutional neural networks for object recognition," in *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2015, pp. 1131–1135.

[11] D. C. Ciresan, U. Meier, J. Masci, L. M. Gambardella, and J. Schmidhuber, "Flexible, high performance convolutional neural networks for image classification," in *Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence*, 2011.

[12] Y. Kawano and K. Yanai, "Food image recognition with deep convolutional features," in *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication*, ser. UbiComp '14 Adjunct. New York, NY, USA: Association for Computing Machinery, 2014, p. 589–593. [Online]. Available: https://doi.org/10.1145/2638728.2641339

[13] M. Szarvas, A. Yoshizawa, M. Yamamoto, and J. Ogata, "Pedestrian detection with convolutional neural networks," in *IEEE Proceedings. Intelligent Vehicles Symposium, 2005.*, 2005, pp. 224–229.

[14] D. Maturana and S. Scherer, "Voxnet: A 3d convolutional neural network for real-time object recognition," in *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2015, pp. 922–928.

[15] S. Zhi, Y. Liu, X. Li, and Y. Guo, "Lightnet: A lightweight 3d convolutional neural network for real-time 3d object recognition," in *Proceedings of the Workshop on 3D Object Retrieval*, ser. 3Dor '17. Goslar, DEU: Eurographics Association, 2017, p. 9–16. [Online]. Available: https://doi.org/10.2312/3dor.20171046

[16] S. Luan, C. Chen, B. Zhang, J. Han, and J. Liu, "Gabor convolutional networks," *IEEE Transactions on Image Processing*, vol. 27, no. 9, pp. 4357–4366, 2018.

[17] C. Liu, W. Ding, X. Wang, and B. Zhang, "Hybrid gabor convolutional networks," *Pattern Recognition Letters*, vol. 116, pp. 164 – 169, 2018.

[18] S. Molaei, M. Shiri, K. Horan, D. Kahrobaei, B. Nallamothu, and K. Najarian, "Deep convolutional neural networks for left ventricle segmentation," in *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2017, pp. 668–671.

[19] D. Gabor, "Theory of communication," *J. Inst. Elec. Eng. (London)*, vol. 93, pp. 429–457, 1946.

[20] J. Daugman, "Uncertainty relation for resolution in space, spatial frequency and orientation optimized by two-dimensional visual cortical filters," *J. Opt. Soc. Amer. A.*, vol. 2, 1985.

[21] A. Kumar and G. K. H. Pang, "Defect detection in textured materials using gabor filters," *IEEE Transactions on Industry Applications*, vol. 38, no. 2, pp. 425–440, 2002.

[22] J. Jing, X. Fang, and P. Li, "Automated fabric defect detection based on multiple gabor filters and kpca," *International Journal of Multimedia and Ubiquitous Engineering*, vol. 11, no. 6, pp. 93–106, 2016.

[23] M. A. El-Sayed, M. Hassaballah, and M. A. Abdel-Latif, "Identity verification of individuals based on retinal features using gabor filters and svm," *Journal of Signal and Information Processing*, vol. 7, 2016.

[24] S. Gornale, A. Patil, and V. C., "Fingerprint based gender identification using discrete wavelet transform and gabor filters," *International Journal of Computer Applications*, vol. 152, no. 4, 2016.

[25] S. T. H. Rizvi, G. Cabodi, P. Gusmao, and G. Francini, "Gabor filter based image representation for object classification," in *2016 International Conference on Control, Decision and Information Technologies (CoDIT)*, 2016, pp. 628–632.

[26] S. Avinash, K. Manjunath, and S. S. Kumar, "An improved image processing analysis for the detection of lung cancer using gabor filters and watershed segmentation technique," in *2016 International Conference on Inventive Computation Technologies (ICICT)*, vol. 3, 2016, pp. 1–6.

[27] A. Daamouche, D. Fares, I. Maalem, and K. Zemmouri, "Unsupervised method for building detection using gabor filters," in *Special issue of the 2nd International Conference on Computational and Experimental Science and Engineering (ICCESEN 2015)*, vol. 130, 2016.

[28] G. Hemalatha and C. P. Sumathi, "Preprocessing techniques of facial image with median and gabor filters," in *2016 International Conference on Information Communication and Embedded Systems (ICICES)*, 2016, pp. 1–6.

[29] S. Lefkovits, L. Lefkovits, and S. Emerich, "Detecting the eye and its openness with gabor filters," in *2017 5th International Symposium on Digital Forensic and Security (ISDFS)*, 2017, pp. 1–5.

[30] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, vol. 1, 2001, pp. I–I.

[31] T. Pumlumchiak and S. Vittayakorn, "Facial expression recognition using local gabor filters and pca plus lda," in *2017 9th International Conference on Information Technology and Electrical Engineering (ICITEE)*, 2017, pp. 1–6.

[32] M. Mahmood, A. Jalal, and H. A. Evans, "Facial expression recognition in image sequences using 1d transform and gabor wavelet transform," in *2018 International Conference on Applied and Engineering Mathematics (ICAEM)*, 2018, pp. 1–6.

[33] C. Low, A. B. Teoh, and C. Ng, "Multi-fold gabor filter convolution descriptor for face recognition," in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2016, pp. 2094–2098.

[34] Z. Lei, M. Pietikäinen, and S. Z. Li, "Learning discriminant face descriptor," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 2, pp. 289–302, 2014.

[35] J. Lu, V. E. Liong, X. Zhou, and J. Zhou, "Learning compact binary face descriptor for face recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 10, pp. 2041–2056, 2015.

[36] R. Nava, B. Escalante-Ramirez, and G. Cristobal, "Texture image retrieval based on log-gabor features," in *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Appplications*, vol. 7441, 2012, pp. 414–421.

[37] X. Liu, J. B. Lao, and J. S. Pang, "Feature point matching based on distinct wavelength phase congruency and log-gabor filters in infrared and visible images," *Sensors*, vol. 19, 2019.

[38] Z. Fan, S. Zhang, J. Mei, and M. Liu, "Recognition of woven fabric based on image processing and gabor filters," in *2017 IEEE 7th Annual International Conference on CYBER Technology in Automation, Control, and Intelligent Systems (CYBER)*, 2017, pp. 996–1000.

[39] G. Srivastava and R. Srivastava, "Salient object detection using background subtraction, gabor filters, objectness and minimum directional backgroundness," *Journal of Visual Communication and Image Representation*, vol. 62, pp. 330 – 339, 2019.

[40] S. H. Khaleefah, S. A. Mostafa, A. Mustapha, and M. F. Nasrudin, "The ideal effect of gabor filters and uniform local binary pattern combinations on deformed scanned paper images," *Journal of King Saud University - Computer and Information Sciences*, 2019.

[41] D. H. Hubel and T. N. Wiesel, "Receptive fields, binocular interaction and functional architecture in the cat's visual cortex," *The Journal of physiology*, vol. 160, no. 1, pp. 106–154, 1962.

[42] Y. Le Cun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel, "Handwritten digit recognition with a backpropagation network," in *Advances in Neural Information Processing Systems*. Morgan Kaufmann, 1990, pp. 396–404.

[43] S. Ruder, "An overview of gradient descent optimization algorithms," 2017.

[44] M. Chacon and P. Rivas, "Face recognition based on human visual perception theories and unsupervised ann," in *State of the Art in Face Recognition*. IntechOpen, 2009.

[45] 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 (CVPR)*, June 2016.

[46] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2015.

[47] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2015.

[48] T. Ishii, R. Nakamura, H. Nakada, Y. Mochizuki, and H. Ishikawa, "Surface object recognition with cnn and svm in landsat 8 images," in *2015 14th IAPR International Conference on Machine Vision Applications (MVA)*, 2015, pp. 341–344.

[49] A. S. Kumar and E. Sherly, "A convolutional neural network for visual object recognition in marine sector," in *2017 2nd International Conference for Convergence in Technology (I2CT)*, 2017, pp. 304–307.

[50] D. P. Sudharshan and S. Raj, "Object recognition in images using convolutional neural network," in *2018 2nd International Conference on Inventive Systems and Control (ICISC)*, 2018, pp. 718–722.

[51] S. A. Zulkeflie, F. A. Fammy, Z. Ibrahim, and N. Sabri, "Evaluation of basic convolutional neural network, alexnet and bag of features for indoor object recognition," *International Journal of Machine Learning and Computing*, vol. 9, 2019.

[52] J. Wang, J. Lu, W. Chen, and X. Wu, "Convolutional neural network for 3d object recognition based on rgb-d dataset," in *2015 IEEE 10th Conference on Industrial Electronics and Applications (ICIEA)*, 2015, pp. 34–39.

[53] M. Schwarz, H. Schulz, and S. Behnke, "Rgb-d object recognition and pose estimation based on pre-trained convolutional neural network features," in *2015 IEEE International Conference on Robotics and Automation (ICRA)*, 2015, pp. 1329–1335.

[54] K. Wu, E. Wu, and G. Kreiman, "Learning scene gist with convolutional neural networks to improve object recognition," in *2018 52nd Annual Conference on Information Sciences and Systems (CISS)*, 2018, pp. 1–6.

[55] S. Hayat, S. Kun, Z. Tengtao, Y. Yu, T. Tu, and Y. Du, "A deep learning framework using convolutional neural network for multi-class object recognition," in *2018 IEEE 3rd International Conference on Image, Vision and Computing (ICIVC)*, 2018, pp. 194–198.

[56] A. Dosovitskiy, J. T. Springenberg, M. Riedmiller, and T. Brox, "Discriminative unsupervised feature learning with convolutional neural networks," in *Advances in Neural Information Processing Systems 27*. Curran Associates, Inc., 2014, pp. 766–774. [Online]. Available: http://papers.nips.cc/paper/5548-discriminative-unsupervised-feature-learning-with-convolutional-neural-networks.pdf

[57] Z. Chen, O. Lam, A. Jacobson, and M. Milford, "Convolutional neural network-based place recognition," 2014.

[58] H. Kataoka, K. Iwata, and Y. Satoh, "Feature evaluation of deep convolutional neural networks for object recognition and detection," 2015.

[59] Y. Cao, Y. Chen, and D. Khosla, "Spiking deep convolutional neural networks for energy-efficient object recognition," *International Journal of Computer Vision*, vol. 113, pp. 54–66, 2015.

[60] W. Fang, L. Ding, B. Zhong, P. E. Love, and H. Luo, "Automated detection of workers and heavy equipment on construction sites: A convolutional neural network approach," *Advanced Engineering Informatics*, vol. 37, pp. 139 – 149, 2018.

[61] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell, "Caffe: Convolutional architecture for fast feature embedding," in *Proceedings of the 22nd ACM International Conference on Multimedia*, ser. MM '14. New York, NY, USA: Association for Computing Machinery, 2014, p. 675–678. [Online]. Available: https://doi.org/10.1145/2647868.2654889

[62] Z. Gao, D. Wang, Y. Xue, G. Xu, H. Zhang, and Y. Wang, "3d object recognition based on pairwise multi-view convolutional neural networks," *Journal of Visual Communication and Image Representation*, vol. 56, pp. 305 – 315, 2018.

[63] B. Ahn, "Real-time video object recognition using convolutional neural network," in *2015 International Joint Conference on Neural Networks (IJCNN)*, 2015, pp. 1–7.

[64] M. Radovic, O. Adarkwa, and Q. Wang, "Object recognition in aerial images using convolutional neural networks," *Journal of Imaging*, vol. 3, 2017.

[65] A. Garcia-Garcia, F. Gomez-Donoso, J. Garcia-Rodriguez, S. Orts-Escolano, M. Cazorla, and J. Azorin-Lopez, "Pointnet: A 3d convolutional neural network for real-time object class recognition," in *2016 International Joint Conference on Neural Networks (IJCNN)*, 2016, pp. 1578–1584.

[66] Jing Huang and Suya You, "Point cloud labeling using 3d convolutional neural network," in *2016 23rd International Conference on Pattern Recognition (ICPR)*, 2016, pp. 2670–2675.

[67] Y.-C. Du, M. Muslikhin, T.-H. Hsieh, and M.-S. Wang, "Stereo vision-based object recognition and manipulation by regions with convolutional neural network," *Electronics 2020*, vol. 9, 2020.

[68] X. Xu, A. Dehghani, D. Corrigan, S. Caulfield, and D. Moloney, "Convolutional neural network for 3d object recognition using volumetric representation," in *2016 First International Workshop on Sensing, Processing and Learning for Intelligent Machines (SPLINE)*, 2016, pp. 1–5.

[69] H. Yao, L. Chuyi, H. Dan, and Y. Weiyu, "Gabor feature based convolutional neural network for object recognition in natural scene," in *2016 3rd International Conference on Information Science and Control Engineering (ICISCE)*, 2016, pp. 386–390.

[70] C. Jiang and J. Su, "Gabor binary layer in convolutional neural networks," in *2018 25th IEEE International Conference on Image Processing (ICIP)*, 2018, pp. 3408–3412.

[71] M. M. Taghi Zadeh, M. Imani, and B. Majidi, "Fast facial emotion recognition using convolutional neural networks and gabor filters," in *2019 5th Conference on Knowledge Based Engineering and Innovation (KBEI)*, 2019, pp. 577–581.

[72] A. Alekseev and A. Bobe, "Gabornet: Gabor filters with learnable parameters in deep convolutional neural network," in *2019 International Conference on Engineering and Telecommunication (EnT)*, 2019, pp. 1–4.

[73] S. Molaei and M. E. Shiri Ahmad Abadi, "Maintaining filter structure: A gabor-based convolutional neural network for image analysis," *Applied Soft Computing*, vol. 88, p. 105960, 2020.