the 2014 ImageNet competition [3]. VGG network has more layers than AlexNet, making it deeper. The winner of that competition was GoogLeNet [4], which is both wider and deeper than AlexNet. This trend of constructing wider and deeper networks to achieve higher accuracy has become prevalent in the field [5], [6], [7]. However, the wider and deeper the network, the more parameters and computational costs it involves. This results in large model size, making real-time classification infeasible, even with powerful Graphics Processing Units (GPUs). In addition, the training time for such wide and deep networks can be prohibitively long due to the sheer number of operations with decimal points required, and the memory requirements during training are substantial due to the high number of parameters.

As the depth of CNNs increases, the vanishing gradient problem arises. This issue occurs when the information about the input in forward propagation and gradients that pass through various layers during backpropagation are lost as they reach the end of the network. To mitigate this problem, researchers have proposed several approaches, including using new activation functions [8] and regularization methods such as dropout. Although these methods can be effective in preventing the vanishing gradient problem, they do not provide a complete solution [9]. Another approach is the concept of skip connections or shortcut connections in networks [10], [11], [12], enabling the network to reuse features from previous layers. Several works have been inspired by the Residual Network (ResNet) [10] model [5], [13], [14].

Another approach for encouraging feature reuse and addressing the vanishing gradient problem is through the use of Densely Connected Convolutional Network (DenseNet) structures [15]. DenseNet addresses this issue by allowing each layer to have access to the collective information from all previous layers within the same block. The current layer in DenseNet receives concatenated feature maps from all previous layers within the same block. The concepts of DenseNet and ResNet have been employed in various visual recognition tasks, including depth estimation, semantic segmentation, instance segmentation, and object detection [16], [17], [18], [19].

DenseNet architectures have consistently delivered robust performance in computer vision tasks, primarily due to their efficient feature reuse. However, these architectures also present several limitations, such as a limited receptive field, challenges with gradient propagation in deep networks, inadequate integration of multi-scale features, and substantial computational costs [20]. The conventional DenseNet architecture attempts to address the limited receptive field by using convolutions and pooling layers, but these measures may not be sufficient for capturing the comprehensive understanding required for certain tasks. Despite the advantages of dense connectivity, very deep networks still face difficulties in gradient propagation, which negatively impacts training convergence. Additionally, DenseNet does not inherently support the integration of multi-scale features, which is crucial for tasks such as object detection and semantic segmentation [21]. While dense connections promote feature reuse, they also introduce significant computational overhead.

To overcome these limitations, we propose an architecture that leverages the core principles of DenseNet and ResNet. Our approach incorporates multi-scale skip connections within the DenseNet framework. This method aims to expand the receptive field by integrating skip connections across multiple scales, potentially capturing a broader context without the need for additional convolutional layers. By explicitly incorporating multi-scale features, our proposed architecture leverages fine-grained and coarse-grained information, crucial for object recognition across various scales. By including multi-scale skip connections, we seek to enhance gradient flow, thereby mitigating vanishing gradient issues that can occur in deep networks. The strategic placement of skip connections reduces redundant computations, improving computational efficiency. These improvements represent a significant advancement in the DenseNet architecture, effectively addressing its inherent limitations and paving the way for more robust neural network designs.

The proposed architecture, as illustrated in Figure 1, embodies a lightweight CNN model that balances computational efficiency and accuracy. Additionally, the structure of the proposed Multi-Scale Skip Connection within the DenseNet context is intuitive and comprehensible. Our aim is to facilitate maximal information flow from the input to the final layers of the network, thereby enhancing overall performance.

To evaluate the performance of the proposed model, we employed four benchmark datasets commonly used in the computer vision community: Canadian Institute For Advanced Research - CIFAR-10 [22], CIFAR-100 [22], Street View House Numbers - SVHN [23], and Imagenette [24]. The model was tested with varying newly introduced feature maps (growth rates ‘‘k’’) and layer depths. Our experimental results demonstrate that our model achieves comparable accuracy to existing networks with significantly fewer floating point operations (FLOPs) calculated in Giga Multiply-Accumulate (GMAC) and additional parameters. Contributions of our research work are:

- We introduce a progressive, straightforward, lightweight, and efficacious Multi-Scale Skip Connection within the context of DenseNet.
- Our solution enables maximal information flow, facilitating the propagation of features throughout the network, thereby mitigating the vanishing gradient problem.
- We conducted a comprehensive evaluation of our proposed model using four benchmark datasets, namely CIFAR-10, CIFAR-100, SVHN, and Imagenette, demonstrating its performance across six distinct network configurations for SVHN, CIFAR-10,