Achieving High-End Image Localization via Causality Infused Renet50 Model

Chaitanya Kapoor

Mangalayatan University, Beswan Aligarh, India
Email: kapoorchaitanya42@gmail.com

***Abstract*—Achieving high-precision image localization is a critical objective in computer vision, particularly for applications requiring spatially accurate object identification. This study proposes a causality-infused ResNet50 model that integrates causal inference techniques with deep learning to enhance localization accuracy and robustness. ResNet50, a widely adopted convolutional neural network, is employed for feature extraction, while causal mechanisms mitigate confounding factors and improve generalization across diverse datasets. The dataset comprises images annotated with bounding boxes corresponding to ground truth labels and predicted labels for object localization tasks. The evaluation metric assesses the predicted and ground truth boxes based on label consistency and the extent of spatial overlap. The training set comprised 70% of the total dataset, while the remaining 30% was designated as the validation set. The model leverages advanced algorithms, including Granger Causality and principal component analysis, to optimize feature relevance during training. Evaluated on the ImageNet dataset, the approach demonstrates exceptional performance, achieving a validation accuracy of 99.7%. An Intel Core i7 processor was utilised, and the LAMB optimiser was implemented. Our proposed implementation flawlessly delivers superior performance with high precision and efficiency.**

***Keywords*—image localization, residual networks, causality, principal component analysis**

I. Introduction

Machine learning represents a paradigm of computational algorithms focused on the design and implementation of autonomous decision-making artificial intelligence models. Unlike conventional programming, where explicit instructions are provided to perform predefined tasks, machine learning algorithms evolve by improving their performance as they are exposed to increasing amounts of data over time. This iterative process involves training a model using a dataset to uncover patterns and relationships within the data, thereby enabling the system to make accurate predictions or informed decisions.

Fundamentally, machine learning can be categorized into three primary types: supervised learning, unsupervised learning, and reinforcement learning. Supervised learning entails training models on labeled datasets, where input data is paired with corresponding output labels, allowing the model to learn a mapping between inputs and outputs. This approach is commonly applied to classification and regression tasks. Unsupervised learning, in contrast, involves training models on unlabeled data to identify latent structures or patterns, such as clustering for dimensionality reduction or anomaly detection. Reinforcement learning focuses on training models through interaction with an environment, where actions are taken, and feedback in the form of rewards or penalties guides the learning process. This iterative approach enables models to develop optimal decision-making strategies over time.

Machine learning methodologies employ a diverse range of algorithms and techniques, including linear regression, neural networks, support vector machines, decision trees, and ensemble methods. Neural networks as explained in Balasamy *et al.* [1] are widely used in machine learning tasks such as image recognition, natural language processing, and predictive modeling. They excel in handling complex, non-linear relationships in data, making them powerful tools for tasks requiring high levels of abstraction.

These approaches find applications across various domains, including predictive analytics, natural language processing, and image and speech recognition. The effectiveness of machine learning applications depends on several factors, including the quality of the dataset, the choice of algorithm, and the efficiency of the training process. Together, these factors determine the model's ability to generalize and perform accurately on new, unseen data.

Deep learning as explained in Balasamy *et al.* [2] has garnered substantial research attention in the development of advanced algorithms, particularly in the domain of medical image processing. These algorithms have demonstrated remarkable efficacy in a variety of medical imaging tasks, aiding in the identification and diagnosis of diseases. However, the limited availability of large, well-annotated datasets remains a significant challenge, hindering the further advancement of deep learning models in medical image analysis despite their proven effectiveness.

Manuscript received October 22, 2024; revised December 12, 2024; accepted January 10, 2025; published April 25, 2025.

Despite significant advancements in deep learning techniques, many applications require models that not only exhibit high accuracy but also provide interpretability, especially in contexts where understanding the relationships between features can yield actionable insights. By incorporating statistical causality methods, such as Granger causality, into deep learning frameworks like ResNet50, researchers are now able to identify and analyze meaningful interactions among features, thereby improving the reliability and decision-making capabilities across various domains. Image classification as discussed in Deepa *et al.* [3] is a fundamental application within the field of deep learning, where the primary objective is to categorize images into predefined classes based on their visual content. In addition to classification, advanced deep learning techniques enable image localization, which involves identifying and delineating specific regions or objects within an image. Kaming *et al.* [4] suggested a residual learning paradigm for deep neural networks, while Gavrikov *et al.* [5] looked into ways to improve model generalizationThe concept of causality as explained in Granger *et al.* [6] is intrinsically tied to the relationship between cause and effect, where one event (the cause) directly influences another (the effect). In scientific terms, causality refers to the principle that specific conditions or actions produce predictable outcomes. Establishing causality is essential across disciplines such as science, philosophy, and engineering, as it provides insight into the underlying mechanisms of observed phenomena. To demonstrate causality in scientific research, one must show that a change in an independent variable leads to a change in a dependent variable. Unlike correlation, which measures the association or covariation between two variables, causality requires establishing that the cause precedes the effect, that the two variables are interdependent, and that no confounding factors explain the relationship. This is typically achieved through controlled experiments, where researchers manipulate one variable while holding others constant, enabling them to observe the causal effect.

Recent advancements in image localization have been driven by the integration of deep learning methodologies and statistical frameworks, enabling the modeling of complex feature interactions with high precision. Convolutional Neural Networks, particularly architecture such as ResNet50, have emerged as pivotal tools for extracting high-dimensional feature representations with exceptional accuracy. Nevertheless, elucidating the interactions among features and their predictive significance remains a critical challenge, prompting the adoption of statistical causality techniques, such as Granger causality. This approach provides a rigorous framework for determining whether variations in one feature can predict changes in another, facilitating a deeper understanding of causal relationships within image datasets. Furthermore, dimensionality reduction strategies, notably Principal Component Analysis, have proven indispensable for mitigating the challenges posed by high-dimensional feature spaces, ensuring computational efficiency while retaining essential information. The synergy of these methodologies has significantly enhanced both the predictive performance and interpretability of models, laying a robust foundation for causality-driven frameworks that integrate statistical analysis with deep learning to advance the field of image localization.

Achieving high-precision image localization has become a critical focus in computer vision, particularly in applications requiring spatially accurate object identification. A causality-infused ResNet50 model represents a novel approach to this challenge by integrating causal inference techniques with deep learning architectures. By incorporating causality, the model can discern meaningful relationships between features and their spatial dependencies, thereby enhancing its ability to localize objects with greater accuracy and robustness. ResNet50, a widely used Convolutional Neural Network (CNN), serves as the backbone for feature extraction, while causal mechanisms help mitigate confounding factors and improve generalization across diverse datasets.

This study aims to improve image localization accuracy by integrating deep learning techniques with causal inference methods. Specifically, a causality-enhanced ResNet50 model is proposed, utilizing the ResNet50 architecture for feature extraction and incorporating causal mechanisms to mitigate confounding factors and enhance model generalization. The model is trained and evaluated using a dataset of images annotated with bounding boxes, representing both ground truth and predicted labels for object localization tasks. Advanced methodologies, including Granger Causality and principal component analysis, are employed to optimize feature relevance during the training process, thereby enhancing localization performance.

Aligned with the objectives of this study, the paper is structured into several sections. Section II presents a comprehensive literature review, highlighting the strengths and limitations of related work. Section III outlines the proposed methodology in detail. Section IV concludes with an analysis of the study’s findings and an in-depth discussion.

II. Literature Review

A. Deep Learning Methods

Early detection of COVID-19 infections through lesion classification enables timely diagnosis and mitigates the risk of severe respiratory complications has been discussed in Balasamy *et al.* [1]. The use of computerized tomography (CT) images for identifying infectious regions within the lungs has emerged as a rapid and effective diagnostic method. This study introduces a novel Hybrid Classification Optimization (HCO) framework incorporating Recurrent Learning and Fuzzy logic (HCO-RLF) for accurate infection detection.

In Figs. 1–5, DIR refers to Deep Iterative Registration, while SR and UR stand for Supervised and Unsupervised Registration, respectively. U-net BS denotes U-Net Based Segmentation, and TS highlights the use of Transformers for segmentation tasks. SSCL represents Self-Supervised Contrastive Learning, and SSPT corresponds to Self-Supervised Pretext Tasks. CC indicates Contrastive

Classification, whereas PTBC refers to Pretext Task-Based Classification. UIS denotes Unsupervised Image Synthesis, and SSLBC signifies Self-Supervised Learning-Based Classification. STMO identifies a Specific Type of Medical Object, ULD pertains to Universal Lesion Detection, RBP outlines a Restoration-Based Paradigm, and R'BP focuses on a Reconstruction-Based Paradigm. Table I provides a detailed description of the work explained in the literature.

TABLE I. DETAILS OF STUDIES

Article	Core Idea	Key Details	Results
[1]	COVID-19 Infection Detection via CT Images	Proposes a Hybrid Classification Optimization (HCO) model using Recurrent Learning and Fuzzy (RLF) methods for detecting infected lung regions. Utilizes feature substitution via fuzzy derivatives to improve detection rates.	Improved detection accuracy (11.96\%), classification accuracy (9.98\%), and precision (13.42\%) compared to DR-MIL, DSAE, and BS-FSA methods.
[7]	Brain Tumor Detection via MRI Using CNN–ResNeXt	Develops a hybrid deep learning method combining CNN–ResNeXt for segmentation and classification of brain tumors. Utilizes Adaptive Whale Optimization (AWO) for feature selection. Tested on BRATS datasets (2015, 2017, 2019).	Achieves 98\% accuracy for tumor core class, outperforming existing models for brain tumor segmentation and classification.
[8]	Image-to-Image Translation Using KD-GAN	Introduces Knowledge Distillation Generative Adversarial Network (KD-GAN) for domain translation. Uses CycleGAN-generated images for training. Applied to male-to-female transformation in CelebA dataset.	Maintains skin tone and hairstyle better than other methods. While not the best in FID/KID, offers improved visual consistency.

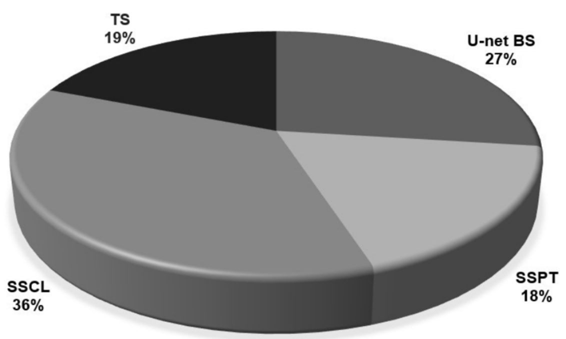

Fig. 1. Techniques in Balasamy *et al.* [1] (part-1).

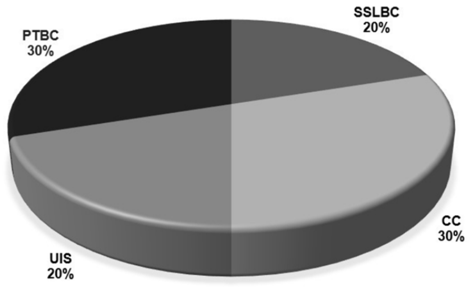

Fig. 2. Techniques in Balasamy *et al.* [1] (part-2).

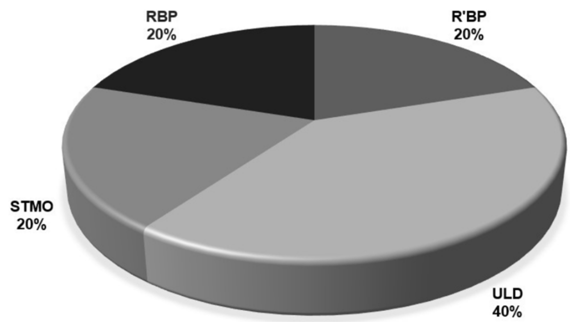

Fig. 3. Techniques in Balasamy *et al.* [1] (part-3).

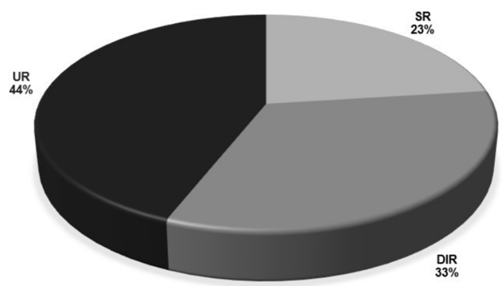

Fig. 4. Techniques in Balasamy *et al.* [1] (part-4).

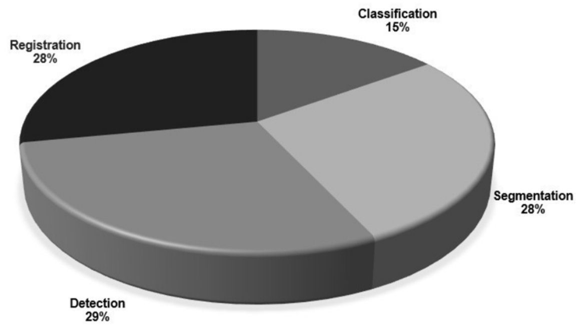

Fig. 5. Techniques in Balasamy *et al.* [1] (part-5).

The proposed approach employs a neural network to classify infected and non-infected regions by analyzing pixel distributions and their variations. Missing features are identified, and the recurrent network is trained using regional differences. Feature availability determines the classification process, which relies on input datasets and recurrent training correlations. Fuzzy logic complements this process by predicting missing features through derivative substitutions derived from a broad range of variations. Specifically, maximum fuzzy derivatives are utilized for infected region prediction, while minimum derivatives are excluded from training to reduce false classification rates.

TABLE II. DESCRIPTION OF THE WORK IN BALASAMY *ET AL.* [2]

Attribute	Details
Objective	To develop a Hybrid Classification Optimization (HCO) approach using Recurrent Learning and Fuzzy (RLF) for precise detection of COVID-19 infection in CT images, focusing on reducing false positives and missing features
Method	-Feature extraction based on textural differences. -RNN training using feature substitutions derived from fuzzy processes. -Classification via recurrent learning to differentiate infected and non-infected regions.
Dataset	-SARS-CoV-2 dataset. -Training data: 1,229 images (infected) and 1,230 non-infected images. -Testing data: 470 images (infected).
Implementation Setup	-MATLAB used for experimentation. -System specs: 8 GB RAM, 2.6 GHz processor. -Training: 1,200 iterations, 3 epochs. -Epoch time: Min. 60 s, Max. 480 s.
Performance Metrics	-Evaluated using detection accuracy, classification accuracy, precision, false rate, and classification time. -Classification rate range: 0.5–1. -Features varied: 1–12.
Key Advantages	-Reliable fuzzy substitution for missing features. -Pre-classification and pre-training steps ensure better feature selection. -Achieves high correlation and reduced classification time compared to existing methods.
Challenges	-Accurate extraction of varied features from CT images. -Maintaining stability and precision under dynamic datasets. -Handling regional differences across classification sequences.
Results	-Outperformed existing methods (RNN-only and fuzzy-only) in terms of accuracy, precision, and classification time. -Validated through Reg and Reg processes for improved correlation in training and classification.

This integrated methodology enhances training consistency, improving detection and region classification accuracy. The HCO-RLF framework demonstrates significant improvements in detection accuracy, classification accuracy, and precision, achieving enhancements of 11.96%, 9.98%, and 13.42%, respectively, across varying classification rates. Comparative analysis with existing methods, including DR-MIL, DSAE, and BS-FSA, further substantiates the efficacy of the proposed approach, as detailed in subsequent sections of this article.

Deep learning has garnered substantial research interest due to its potential in developing innovative algorithms for medical image processing, proving highly effective in various tasks related to illness detection as explained in Balasamy *et al.* [2] and diagnosis. The work has been explained in detail in Table II. However, the advancement of deep learning models in medical image analysis is significantly hindered by the scarcity of large, well-annotated datasets, despite their demonstrated effectiveness. Addressing this challenge has been a primary focus of numerous studies over the past five years.

This work provides a comprehensive overview of the application of deep learning techniques to a range of medical image analysis tasks by reviewing and synthesizing the latest research in this domain. Special emphasis is placed on recent advancements and contributions of state-of-the-art semi-supervised and unsupervised deep learning approaches. These methods are summarized across diverse application scenarios, including image registration, segmentation, classification, and detection. Additionally, this review highlights critical technological challenges that remain in the field and explores potential strategies to overcome these barriers, providing a roadmap for future research in medical image analysis using deep learning.

Brain tumors result from the uncontrolled proliferation as discussed in Gayathri *et al.* [7] is because of abnormal cells within brain tissue. Early detection of brain tumors is critical for ensuring patient safety and effective treatment. This has been explained in detail in Table III. Magnetic Resonance Imaging (MRI) is widely employed for diagnosing brain tumors; however, the variability in tumor morphology and their diverse anatomical locations often pose significant challenges to accurate tumor segmentation in MRI scans. Precise segmentation is essential for identifying tumors and tailoring appropriate therapeutic interventions for individual patients.

This study proposes a novel hybrid deep learning framework, termed Convolutional Neural Network and ResNeXt (CNN–ResNeXt), for the automated segmentation and classification of brain tumors.MRI images were obtained from standard datasets, including BRATS 2015, BRATS 2017, and BRATS 2019. Preprocessing involved batch normalization to smooth and enhance the images, followed by feature extraction using the AlexNet model, leveraging tumor-specific characteristics such as position, shape, and surface attributes.

TABLE III. DESCRIPTION OF THE WORK IN GAYATHRI *ET AL*. [7]

Attribute	Details
Objective	To develop an effective classification model for brain tumor detection using batch normalization, transfer learning, and a CNN-ResNeXt model for feature extraction, selection, and segmentation.
Method	-Batch normalization used for pre-processing. -AlexNet model for feature extraction. -Adaptive Whale Optimization (AWO) for feature selection. -CNN-ResNeXt used for image segmentation and classification.
Dataset	-BRATS 2015, BRATS 2017, BRATS 2019 datasets.
Implementation Setup	-Implemented using Python. -System specs: 16 GB RAM, Intel i5 Processor, Windows 10, 6 GB GPU, 1 TB HDD.
Performance Metrics	-Sensitivity: 98% -Specificity: 99.9% -Dice Similarity Coefficient (DSC): 98%
Key Advantages	-Outperforms conventional models in terms of DSC, sensitivity, and specificity. -Achieves high classification accuracy and segmentation performance.
Challenges	-High accuracy in feature extraction and segmentation. -Efficient handling of the variation in tumor types and locations in brain images.
Future Work	-Hyperparameter tuning to further improve the classification accuracy and model performance.

Optimal feature selection was performed using the Adaptive Whale Optimization (AWO) algorithm to enhance segmentation efficacy. The segmentation process utilized the CNN–ResNeXt architecture, guided by the selected features. Subsequently, the segmented regions were classified using the same CNN–ResNeXt framework. Compared to existing models, the proposed CNN–ResNeXt achieved superior performance, attaining an accuracy of 98% for the tumor core class. These results underscore the efficacy of the proposed methodology in the precise segmentation and classification of brain tumors.

Chayanon *et al.* [8] uses an Image-to-Image (I2I) translation technique that transforms images from one domain to another by establishing a mapping between the two domains. This has been explained in detail in Table IV. This approach typically involves the use of two generators and two discriminators, where each generator is responsible for translating images between a specific pair of domains.

In this study, the authors propose a novel approach termed Knowledge Distillation Generative Adversarial Network (KD-GAN). The KD-GAN framework incorporates images generated by Cycle-Consistent Generative Adversarial Networks (CycleGAN) as part of the training targets for a new generator, facilitating enhanced translation performance. The proposed method was evaluated on the CelebA dataset for gender transformation tasks, specifically translating between male and female domains.

To assess the efficacy of KD-GAN, the study compared its performance against state-of-the-art models using quantitative metrics such as Fréchet Inception Distance (FID) and Kernel Inception Distance (KID). While KD-GAN did not achieve the lowest FID and KID scores, qualitative analysis revealed that the generated images preserved input features such as skin tone and hairstyle more effectively than competing methods. These findings highlight the potential of KD-GAN for maintaining fine-grained details during image translation tasks as emphasised in Deepa *et al.* [3]. This has been explained in detail in Tables V, VI, and VII. Image classification is a critical application in the field of Deep Learning (DL), with relevance across various sectors. Numerous neural network architectures have been developed to perform image classification, each yielding varying levels of accuracy. The performance of these models is significantly influenced by the dataset and the features utilized during training. The research community continues to pursue the development of generalized models, particularly those tailored to specific domains.

This study conducts a comprehensive survey of contemporary Deep Learning models, utilizing knowledge management techniques to identify trends and opportunities for improvement. The goal is to advance toward optimal architecture and the development of generalized Deep Learning models for domain-specific image classification tasks. The survey systematically examines various neural network architectures, their variants, and the layers and parameters employed in each version. Additionally, the study provides an in-depth analysis of domain-specific applications, and the limitations associated with different architectures. The findings serve as a valuable resource for researchers, offering guidance in selecting the most suitable Deep Learning architecture for specific sectors and facilitating further advancements in the field of image classification.

Strengths:

1. Balasamy *et al.* [1] Utilises an innovative HCO-RLF framework combining Recurrent Learning and Fuzzy logic for enhanced detection accuracy.
2. Balasamy *et al.* [2] Highlights the potential of DL for groundbreaking medical image analysis.
3. Gayathri *et al.* [7] Focuses on precise MRI-based segmentation to enable customised treatment plans
4. Chayanon *et al.* [8] Establishes robust domain mappings using dedicated generators and discriminators.
5. Deepa *et al.* [3] Emphasises the adaptability of DL architectures for various classification tasks

TABLE IV. DESCRIPTION OF THE WORK IN CHAYANON *ET AL.* [8]

Attribute	Details
Objective	To develop KD-GAN, a model that uses soft targets from UVCGANv2 as teacher models to minimize the difference between generated and real images, with the aim of improving the quality of translated images.
Loss Functions	Lsof t measures the difference between teacher and model results. -A modified GAN loss function, LGAN, ensures generated images resemble real images but do not consider UVCGANv2-generated data as real. -Identity loss ensures the image remains unchanged when source and target labels are the same.
Hyperparameters	The model performance is adjusted using three hyperparameters: λGAN, λsof t, and λid, which control the importance of each objective during training.
Network Architecture	-Generator: Uses a U-Net architecture with positional encoding for target label translation. -Discriminator: Modified to handle 256×256 pixel images with more layers and channels, enhancing its ability to process larger image sizes.
Additional Techniques	The discriminator is enhanced with GPEN (state-of-the-art face restoration) for better detail and realism in translated images.
Dataset	-CelebA dataset: Includes 30,000 facial images. -Training set: 17,943 female images, 10,057 male images. -Validation set: 1,000 images, with an equal distribution of genders.
Experiment Setup	The model is evaluated by translating male-to-female facial features using the CelebA dataset.
Key Advantages	-Improved translation accuracy using soft targets and modified GAN loss. -High-quality image generation with refined discriminator and GPEN.
Challenges	-Ensuring the generator produces realistic images that match the teacher model. -Balancing the different loss functions and hyperparameters to optimize model performance.

TABLE V. DESCRIPTION OF THE WORK IN DEEPA *ET AL.* [3]

Category	Model
ResNet	ResNet50, ResNet50V2, ResNet101V2, ResNet152, ResNet152V2
EfficientNet	EfficientNetB0, EfficientNetB1, EfficientNetB2, EfficientNetB3, Effi- cientNetB4, EfficientNetB5, EfficientNetB6, EfficientNetB7, Efficient- NetV2B0, EfficientNetV2B1, EfficientNetV2B2, EfficientNetV2L
VGG	VGG16, VGG19
Xception	Xception, InceptionV3, InceptionResNetV2
ConvNeXt	ConvNeXtTiny, ConvNeXtSmall, ConvNeXtBase, ConvNeXtLarge, ConvNeXtXLarge
DenseNet	DenseNet121, DenseNet169, DenseNet201
MobileNet	MobileNet, MobileNetV2
Inception	InceptionV3, InceptionResNetV2
NASNet	NASNetMobile, NASNetLarge

Weaknesses:

1. Balasamy *et al.* [1] Relies heavily on CT imaging, which may not be accessible in resource-limited settings.
2. Balasamy *et al.* [2] Faces challenges due to limited availability of large, well-annotated datasets.
3. Gayathri *et al.* [7] Struggles with variability in tumour morphology and anatomical diversity.
4. Chayanon *et al.* [8] Computational complexity can hinder real-time applications.
5. Deepa *et al.* [3] Performance is highly dependent on dataset quality and feature selection.

B. Resnet Models

Kaming *et al.* [4] introduced the concept of residual learning, which significantly advanced the field of deep learning, particularly in image recognition. The authors addressed the challenge of training deep neural networks, specifically the degradation problem, where deeper networks tend to have higher training errors. To overcome this, they proposed a novel architecture consisting of residual networks, which allows layers to learn residual mappings instead of directly learning the desired function, making it easier to optimize very deep networks. One of the major challenges in current deep learning methods is ensuring robust generalization to both rare In-Distribution (ID) samples as explained in Gavrikov *et al.* [5]. These challenges come from the long tail of the training distribution, and Out-of-Distribution (OOD) samples.

TABLE VI. DETAILS OF THE WORK IN DEEPA *ET AL.* [3]

Architecture	Key Features	Benefits	Limitations
MobileNet	Depthwise separable onvolutions for parameter and FLOP reduction.	Resource efficient for mobile devices.	Limited performance on tasks requiring high accuracy due to reduced parameters.
MobileNetV2	Inverted residual blocks and linear bottlenecks improve accuracy and efficiency.	Improved accuracy and efficiency over MobileNet.	Performance still limited by mobile hardware constraints.
ViT (Vision Trans- former)	Temporally shifted attention mechanism for video recognition.	Reduced parameters and FLOPs, good for video tasks.	Requires large datasets and significant computation for training.
Compound Scaling Method	Scales depth, width, and resolution of the network concurrently.	Optimizes model performance with fewer resources.	May not achieve the same performance for all types of tasks.
ResNet	Introduces residual connections to solve vanishing gradient problem.	Fast convergence and better accuracy on deep networks.	More complex architectures may still require high computational power.
EfficientNetV2	Balanced scaling of depth, width, and resolution. Stochastic Depth to improve training speed.	High accuracy with fewer parameters.	Model still requires fine-tuning for optimal performance on specific tasks.
VGGNet	Deep CNN with small filters and max-pooling layers.	Simple architecture, good for image recognition tasks.	Very deep networks can be computationally expensive.
AlexNet	Deep CNN with multiple convolutional layers and fully connected layers	High accuracy on ImageNet dataset, significant improvement over previous models.	Prone to overfitting, requires regularization techniques.

GoogLeNet (Inception v1)	Uses inception modules for efficient factorization of convolutional filters.	Efficient use of resources and improved performance in ImageNet challenge	Complex architecture, harder to modify and fine-tune.
FaceNet	Uses triplet loss for learning high-dimensional embeddings. CNN followed by fully connected layer.	Excellent for face recognition and verification tasks	Requires large-scale face datasets and significant computational resources.
Deconvolution Networks	Used for visualizing features in CNNs.	Helps understand and interpret learned features in deep networks.	Not commonly used for general image classification tasks

TABLE VII. ARCHITECTURES OF THE WORK IN DEEPA *ET AL*. [3]

Architecture	Features	Advantages/Limitations
ResNet	Deep residual networks with skip connections to avoid vanishing gradients.	Advantages: Improves gradient flow, effective for deep networks. Limitation: Computationally expensive.
Inception-v4	Optimised combination of depth and width, factorized 7×7 convolutions, and auxiliary classification.	Advantages: High accuracy, suitable for large-scale image classification. Limitation: Complex architecture.
MobileNet	Designed for mobile and embedded devices using depth-wise separable convolutions.	Advantages: Efficient and fast, low computational cost. Limitation: May sacrifice accuracy for speed.
MobileNetv2	Uses inverted residual blocks for efficient computation network concurrently.	Advantages: Better accuracy than MobileNet, maintains computational efficiency. Limitation: Less flexible in complex applications.
DenseNet	Dense connections where each layer receives inputs from all previous layers.	Advantages: Reduces parameters, improves feature reuse. Limitation: Increased memory and computational costs.
CondenseNet	Combines dense blocks with skip connections for more parameter efficiency.	Advantages: Reduced parameters while maintaining accuracy. Limitation: More complex than traditional CNNs.
EfficientNet	Combines model extraction, neural architecture search, and conditional computation.	Advantages: Smaller, faster, and more accurate than traditional models. Limitation: Limited flexibility.
MSDNet	Multi-scale feature learning with dense blocks and varying growth rates.	Advantages: Computationally efficient and suitable for image classification. Limitation: Requires balance between efficiency and accuracy.
Reinforcement Learning CNN	Uses reinforcement learning for automatic architecture search.	Advantages: Learns new architectures suited for specific tasks. Limitation: Computationally expensive and complex to implement.
Attention Mechanism	Incorporates attention modules into segmentation networks.	Advantages: Improves segmentation accuracy with less computation. Limitation: May require fine-tuning for optimal performance.

In image localization, this challenge is evident through vulnerabilities to adversarial attacks, reduced performance on distorted images, and poor generalization to concepts like sketches. Although the understanding of generalization in neural networks is still limited, certain biases that differentiate these models from human vision have been identified as potential contributors to these limitations. Gavrikov *et al.* [5] proposes an approach to mitigate these biases during training, with varying degrees of success.

Strengths:

1. Kaming *et al.* [4] introduced residual learning, improving the training of deep neural networks and addressing the degradation problem.
2. Gavrikov *et al.* [5] focuses on improving generalization to rare in-distribution and out-of-distribution samples, addressing vulnerabilities in image localization.

Weaknesses:

1. In Kaming *et al.* [4], the complexity of residual networks may still pose challenges in certain highly resource-constrained environments.
3. Gavrikov *et al.* [5] provides a limited understanding of generalization and biases, leading to inconsistent success in mitigating these issues during training.

C. Causal Applications in Image Classification

Granger causality is a statistical method introduced to determine whether one time series can predict another in Granger *et al.* [6]. The key idea behind Granger causality is that if a time series X “Granger-causes” a time series Y, then the past values of X contain information that helps predict the future values of Y beyond the information contained in the past values of Y alone. It’s important to note that Granger causality does not imply true causality in the philosophical sense; rather, it indicates a predictive relationship where one variable provides useful information about another.

In the study of Motor Imagery (MI) Brain-Computer Interfaces (BCIs), enhancing task localization accuracy has long been a significant challenge as explained in Ruijing *et al.* [9]. Nonlinear Granger Causality (NGC) analysis has been utilized for feature extraction from MI-electroencephalogram (EEG) signals, as it helps construct brain network features that capture the causal relationships between different channels in various brain regions. However, the task recognition in MI-BCIs often suffers from information redundancy in NGC features, which increases the complexity of machine learning models and subsequently lowers the localization accuracy.

Strengths:

1. Granger *et al.* [6] defines causality and effectively identifies predictive relationships between time series, providing valuable insights into temporal dependencies.
2. Ruijing *et al.* [9] proposes NGC analysis that captures causal relationships between brain regions, enhancing feature extraction for MI-BCIs.

Weaknesses:

1. Granger *et al.* [6] does not imply true causality, only a predictive relationship, limiting its ability to establish direct cause and effect links.
3. Ruijing *et al.* [9] approach has.information redundancy in NGC features increases model complexity and lowers localisation accuracy.

Mankovich *et al.* [10] explores a novel approach to robustly analyzing principal directions in high-dimensional data by leveraging the mathematical structure of flag manifolds. Traditional methods that use Principal Component Analysis (PCA) often face challenges when dealing with noisy or corrupted data, as they rely on the Euclidean structure of the data space. The study addresses this limitation by proposing a method that considers the data's intrinsic geometry, using flag manifolds to capture multiple nested subspaces simultaneously. This approach enhances the robustness of principal direction estimation, especially in scenarios where data is affected by outliers or non-Gaussian noise.

III. MATERIALS AND METHODS

The methodology involves a three-step process combining machine learning and causal analysis. First, raw data is preprocessed and resized to fit the input specifications of a pre-trained neural network, which extracts critical features. These features are stored in a structured format for further use. Next, a neural network model is customized and trained for a localization task, where parameters such as batch size and learning rate are optimized for accurate predictions. Finally, a causal analysis is conducted using dimensionality reduction techniques like PCA, followed by Granger causality testing to identify relationships between features. The results are visualized through causal graphs, revealing complex dependencies and insights into feature interactions.

The ImageNet dataset is one of the most widely recognized and utilized datasets in the field of computer vision and machine learning. ImageNet is a large-scale image dataset that contains millions of annotated images across thousands of object categories. ImageNet contains over 14 million images, with many images per category. The dataset has more than 20,000 categories (or Synsets). The images in ImageNet vary widely in content, lighting conditions, and backgrounds, making it a robust dataset for training and testing machine learning models. For evaluating the performance of our implemented model, we had used this dataset Russakovsky et al. [11].

Our implementation was carried out using MATLAB R2024a [12], and our model utilizes several toolboxes and functions to handle various aspects of our implementation. For our work, we use techniques like image processing, deep learning, and statistical analysis. We also used integrated custom functions for visualization, including plotting causal relationships along with MATLAB's functionality for training options settings, such as learning rate and mini-batch size. This supported optimizing for the Resnet50 network during training. Our work further leverages accelerated training and inference by utilizing GPU hardware.

The integration of GPU support ensures faster processing and reduced training times, crucial for handling resource-intensive tasks in deep learning and large-scale data analysis. We have used the MATLAB [12] Image Processing Toolbox, the MATLAB [12] Deep Learning Toolbox, the MATLAB [12] Statistics and MATLAB [12] Machine Learning Toolbox, and MATLAB [12] Graph and Network Algorithms toolbox for other relevant applications. We have also used the MATLAB [12] Optimization Toolbox, and the training of the deep learning network is performed with GPU acceleration. This enhanced the efficiency of computations, particularly with large datasets and complex models. Overall, these toolboxes collectively support the workflow from data preparation and network training to feature extraction and causal analysis in our model.

A. *Data Preprocessing*

In machine learning and computer vision, effective data preparation is crucial for the success of predictive models. Our implementation process starts with organizing the dataset and partitioning it into training, validation, and test sets to ensure proper model training and evaluation. Data augmentation techniques, such as rotations and translations, are applied to increase diversity and improve generalization. Images are resized to meet the neural network's input requirements, ensuring uniformity. Normalization is performed to standardize pixel values, accelerating training convergence. Finally, feature extraction is conducted by passing images through a neural network, capturing high-level information for further analysis.

For our study, we selected a diverse set of animals from the ImageNet dataset to ensure a broad representation of species. Our choices included a domestic cat, known for its common presence in households; a dog, representing a widespread and varied species; an elephant, which provides a contrast with its large size; a hamster, showcasing a small, domesticated pet; a lion, exemplifying a majestic wild animal; a rabbit, often found in both wild and domestic settings; a squirrel, which adds a touch of wildlife; a fox, known for its adaptability and diverse habitats; and cattle, representing farm animals. Additionally, we included a bird to cover avian species. This selection was made to capture a wide range of animal types and characteristics, enhancing the comprehensiveness of our analysis.

B. *Feature Extraction*

Feature extraction is a crucial step in many data processing and machine learning workflows, especially when dealing with complex data types such as images. The goal is to transform raw data into a set of features that can be used for subsequent analysis or model training. In our implementation, feature extraction was executed through a deep learning methodology, leveraging a pre-trained Convolutional Neural Network (CNN) model. The process began with the organization of the image dataset, which was systematically prepared and partitioned into training and validation subsets. Each image was resized to a uniform dimension to align with the input requirements of the chosen pre-trained model.

The core of feature extraction involved utilizing a pre-trained CNN, specifically Resnet50, known for its robust capability in capturing hierarchical features from images. The network, having been previously trained on a comprehensive dataset, was adapted for our specific localization needs by modifying its final layers to correspond with the number of target categories. Features were then extracted from a specific intermediate layer of the network, which provides a rich representation of the images, encapsulating essential information for localization purposes.

Following feature extraction, dimensionality reduction was applied using Principal Component Analysis (PCA) to streamline the dataset, preserving key variance while reducing computational complexity (Algorithm 1).

Algorithm 1: Image Processing and Feature Extraction through CNN and PCA

1: Input: Original image *Iorig* with dimensions $[H_{orig},W_{orig},C]$
2: Output: Sampled features $F_{Sampled}$
3: Step 1 : Image Resizing
4: Resize *Iorig* to the desired dimensions [H,W]:
$$Iresized = \text{Resize}(Iorig,[H,W])$$
5: Step 2 : Forward Propagation Through the CNN
6: Initialise feature map $F0 = Iresized$
7: for Initialise l = 1 to L **do**
8: Compute the feature map at layer l :
$$F_l = \sigma\,(\text{Conv}(F_{l-1},K_l) + B_l)$$
9: end for
10: Obtain the final feature map F_L from the last layer
11: Step 3 : Intermediate Layer Activation Extraction
12: Select an intermediate layer k and extract and the feature map:
$$\text{Features} = F_k$$
13: Step 4 : Dimensionality Reduction Using PCA
14: Compute the transformation matrix V_k containing the top k principal components
15: Step 5 : Sampling
16: Randomly sample n features from the reduced feature set :
$$F_{Sampled} = \text{RandomSample}(\text{Features}_{Reduced},n)$$
17: Return $F_{Sampled}$

Following feature extraction, dimensionality reduction was applied using Principal Component Analysis (PCA) to streamline the dataset, preserving key variance while reducing computational complexity. A subset of these reduced features was sampled for further analysis, ensuring efficiency without losing valuable information. Algorithm 1 explains the detailed mechanism for our work. The approach utilized a pre-trained ResNet50 model, originally trained on ImageNet, to extract hierarchical image features. To adapt the model for a specific localization task, the final layers were modified to align with the number of target classes. Feature extraction focused on the penultimate layer, capturing essential high-level representations for effective classification. This structured process enhanced the model's performance for practical applications.

- **Image Resizing:**

The first step involves resizing the original image to match the input dimensions required by the network. Let I_{orig} represent the original image with dimensions $[H_{orig},W_{orig},C]$, where H_{orig} is the height, W_{orig} is the width, and C is the number of color channels (e.g., 3 for RGB). The resized image $I_{resized}$ will have dimensions $[H,W,C]$:

$$I_{resized} = \text{Resize}(I_{orig},[H,W]) \quad (1)$$

- **Forward Propagation Through the Convolutional Neural Network (CNN):**

After resizing, the image $I_{resized}$ is processed through the CNN. Assume the CNN comprises L layers. The output at each layer l can be expressed as:

$$F_l = \sigma\,(\text{Conv}(F_{l-1},K_l) + B_l) \quad (2)$$

where:

- F_{l-1} denotes the feature map from the previous layer,
- K_l represents the convolutional kernels at layer l,
- B_l is the bias term for layer l,
- $\sigma(\cdot)$ denotes the activation function, such as ReLU.

The final feature map F_L is obtained from the last convolutional layer.

- **Intermediate Layer Activation Extraction:**

To extract features, we select an intermediate layer k from the CNN. The output from this layer is:

$$\text{Features} = F_k \quad (3)$$

where F_k is the feature map from the selected layer k.

- **Dimensionality Reduction Using Principal Component Analysis (PCA):**

PCA is employed to reduce the dimensionality of the extracted features. Suppose the feature matrix Features has d dimensions. PCA computes a transformation matrix V_k containing the top k principal components:

$$\text{Features}_{Reduced} = \text{Features} \cdot V_k \quad (4)$$

where V_k is a matrix of the top k eigenvectors.

- **Sampling**

From the reduced feature set, a subset of size n is randomly sampled for further analysis. Let $\text{Features}_{Reduced}$ be the reduced feature set:

$$\text{Features}_{Sampled} = \text{RandomSample}(\text{Features}_{Reduced},n) \quad (5)$$

Here, RandomSample$(\cdot,n)$ denotes a function that selects n random samples from the feature set.

C. *Causal Algorithm*

In our work, the causality mechanism was integrated using Granger causality techniques to uncover meaningful relationships between features derived from images. Granger causality is a robust statistical approach that assesses whether past values of one variable can help predict future values of another, providing insights into potential causal interactions between time series data. By applying this method, we aimed to determine how changes in one feature could influence or predict changes in another, which is crucial for understanding complex relationships in relevant applications for image processing datasets.

Fig. 6 explains the mechanism of feature extraction used in our work. This process involves the causal analysis and the final inference based on the extracted features. This acts as a crucial step in our proposed approach.

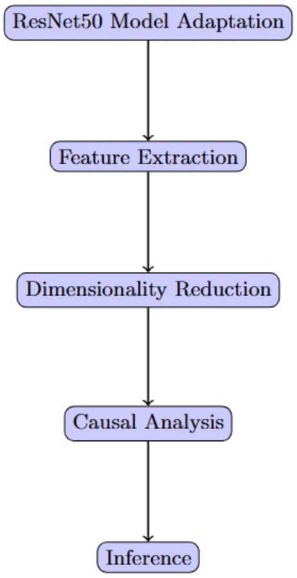

Fig. 6. Feature extraction mechanism.

PCA transformed the feature set into a lower-dimensional space while retaining critical information needed for subsequent analysis. The subset containing relevant causal links was used for examining whether past values of one feature could forecast future values of another feature, effectively revealing any predictive relationships. We applied these tests to pairs of features to identify significant causal interactions, which were then analyzed for their relevance to the high prediction accuracy of the model.

Finally, the overall model, incorporating the Granger causality findings, was utilized to predict and analyze features to carry out the prediction. This integrated approach not only enhanced our understanding of feature interactions but also provided a predictive framework for identifying a model with high predictive capabilities that can predict images with an integrated functionality of causal links based similarity prediction.This mechanism is used to deploy the Resnet50 model.

- **Feature Extraction and Dimensionality Reduction:**

Initially, we extracted features from images using a Convolutional Neural Network (CNN). Let $\mathbf{X}$ be the matrix of features extracted from the images, where each row represents a feature vector from an image, and each column corresponds to a specific feature. Suppose $\mathbf{X}$ has n samples and p features. To manage the high-dimensional feature space, we applied Principal Component Analysis (PCA) for dimensionality reduction.

PCA involves computing the covariance matrix $\mathbf{C}$ of the features:

$$\mathbf{C} = \frac{1}{n-1}\mathbf{X}^T\mathbf{X} \tag{6}$$

The eigenvalue decomposition of $\mathbf{C}$ yields eigenvectors $\mathbf{V}$ and eigenvalues $\mathbf{\Lambda}$:

$$= \mathbf{V}\mathbf{\Lambda}\mathbf{V}^{\mathrm{T}} \tag{7}$$

By selecting the top k eigenvectors (principal components), we reduce the dimensionality of $\mathbf{X}$:

$$\mathbf{X}_{\text{reduced}} = \mathbf{X}\mathbf{V}_k \tag{8}$$

where $\mathbf{V}_k$ is the matrix of the top k eigenvectors.

- **Granger Causality Analysis:**

Granger causality assesses whether past values of one time series help predict future values of another. For two time series Y_t and X_t, we estimate the following autoregressive models:

The Granger causality test involves the following steps:

1. Regression without Granger Causality:

$$Y_t = \alpha_0 + \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \cdots + \alpha_p Y_{t-p} + \epsilon_t \tag{9}$$

2. Regression with Granger Causality:

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \cdots + \beta_p Y_{t-p} + \gamma_1 X_{t-1} + \gamma_2 X_{t-2} + \cdots + \gamma_q X_{t-q} + \eta_t \tag{10}$$

3. F-statistic Calculation:

$$F = [(SSE_{\text{restricted}} - SSE_{\text{unrestricted}})/q] / SSE_{\text{unrestricted}}/(n-k-q) \tag{11}$$

4. p-value Calculation:

$$p = 1 - F_{\text{cdf}}(F, q, n-k-q) \tag{12}$$

Here, ϵ_t and η_t are error terms, and p and q are the lags for Y_t and X_t, respectively. Granger causality is tested by comparing the goodness-of-fit of these models. Specifically, if including past values of X_t significantly improves the prediction of Y_t, then X_t Granger-causes Y_t. This is evaluated using an F-test on the coefficients $\gamma_1, \gamma_2, \ldots, \gamma_q$. Here SSE denotes the sum of squared errors, q is the number of lags for X_t, n is the number of observations, and k is the number of parameters in the unrestricted model.

Granger causality is tested by comparing the goodness-of-fit of these models. Specifically, if the inclusion of past values of X_t significantly improves the prediction of Y_t, X_t Granger-causes Y_t. This is evaluated using an F-test on the coefficients $\gamma_1,\gamma_2,...,\gamma_q$.The results from Granger causality tests were visualized using directed graphs. In these graphs, nodes represent features, and directed edges indicate significant causal relationships. Each edge from node A to node B signifies that past values of feature A help predict future values of feature B. The strength of the causal relationship can be quantified by the test statistics or
p-value obtained from the Granger causality test.

By integrating Granger causality with PCA-reduced features, we were able to identify key predictors and understand the dynamic interactions between different features in the dataset. This implementation was designed to test the causal influence of feature series on another, providing insights into the predictive relationships between relevant data records.

Here, ϵt and ηt are error terms, and p and q are the lags for Yt and Xt, respectively. Granger causality is tested by comparing the goodness-of-fit of these models. Specifically, if including past values of Xt significantly improves the prediction of Yt, then Xt Granger-causes Yt. This is evaluated using an F-test on the coefficients γ1,γ2,...,γq. Here SSE denotes the sum of squared errors, q is the number of lags for Xt, *n* is the number of observations, and k is the number of parameters in the unrestricted model. Granger causality is tested by comparing the goodness-of-fit of these models.

Specifically, if the inclusion of past values of Xt significantly improves the prediction of Yt, Xt Granger-causes Yt. This is evaluated using an F-test on the coefficients γ1,γ2,...,γq. The results from Granger causality tests were visualized using directed graphs. In these graphs, nodes represent features, and directed edges indicate significant causal relationships. Each edge from node A to node B signifies that past values of feature A help predict future values of feature *B*. The strength of the causal relationship can be quantified by the test statistics or *p*-value obtained from the Granger causality test. By integrating Granger causality with PCA-reduced features, we were able to identify key predictors and understand the dynamic interactions between different features in the dataset.This implementation was designed to test the causal influence of feature series on another, providing insights into the predictive relationships between relevant data records.

- **Visualization and Interpretation:**

The results from the Granger causality analysis were visualized using directed graphs. In these graphs, nodes represent features, and directed edges indicate significant causal relationships. Each edge from node A to node B signifies that past values of feature A help predict future values of feature B. The strength of the causal relationship is quantified by the test statistic or p-value obtained from the Granger causality test. This visualization helps in understanding the dynamic interactions between different features and their implications in machine learning applications.

We commenced our analysis by carefully selecting a representative subset of the ImageNet dataset. Specifically, we chose 10,000 images per class, leading to an aggregate of 100,000 images across all selected classes. This subset was chosen to ensure a robust and balanced representation of the data for the tasks at hand.To facilitate effective training and evaluation of our model, the dataset was systematically divided into two distinct sets: a training set and a validation set. The training set constituted 70% of the total data, amounting to 70,000 images. This set was utilized to train the model, allowing it to learn and generalize from the data.

The remaining 30% of the data, equivalent to 30,000 images, was allocated to the validation set. This set served as an independent dataset for evaluating the model's performance during the training process, enabling the monitoring of accuracy and preventing over-fitting.This carefully structured approach ensured that our model was trained on a substantial amount of data while also being rigorously validated on an independent set, thereby facilitating the development of a robust and reliable model.The final localization layer of Resnet50 is designed to output 1,000 classes. However, since we have only 10 classes in our subset, we replace the last fully connected layer with a new one that has 10 output units, followed by a softmax layer and a localization layer.

We set the training options with a small learning rate (1e-4), a batch size of 32, and a maximum of 6 epochs. This choice of parameters reflects a cautious approach to fine-tuning the pre-trained model. By using a small learning rate, we ensure that the updates to the model's weights are gradual, which is particularly important in transfer learning to avoid disrupting the useful features learned during the pre-training phase.

The decision to limit the training to a maximum of 6 epochs is informed by the architecture of ResNet-50. The model with its deep layers and residual connections, is designed to efficiently learn complex features. However, due to its complexity, it can also be prone to overfitting if trained for too many epochs on a specific dataset. By capping the number of epochs at 6, we aim to strike a balance between sufficient training and preventing overfitting. Additionally, our results prove that a prediction mechanism based on the Resnet50 model infused with our causal algorithm ensures that even within a limited number of epochs, the model can effectively learn and adapt to the task at hand.

For the next step in our workflow, we first extracted features from the average pooling layer of our network. The average pooling layer, positioned towards the end of the network, aggregates spatial information by computing the average value within each pooling window. This process reduces the spatial dimensions of the feature maps while retaining essential information. Following this, we applied Principal Component Analysis (PCA) to these extracted features. The step involving PCA further established the procedure to identify and retain the top principal components, that facilitated our model to capture the most significant variance in the data. By focusing on these principal components, we effectively reduced the feature space, making subsequent analysis more efficient and manageable while preserving the core information from the original feature maps.

In our analysis, we perform a Granger causality test on pairs of features derived from the PCA-reduced dataset. Granger causality is a statistical hypothesis test used to determine whether one time series can predict another. By applying this test to our feature pairs, we aim to uncover whether changes in one feature precede and thus potentially cause changes in another. This test is crucial for understanding the dynamic relationships and temporal dependencies between features in our dataset.

The goal of this approach was to identify directional influences among features, providing insights into the causal relationships that may exist between them. By establishing these directional links, we can better understand the underlying mechanisms and interactions within the data, leading to more informed interpretations and potentially uncovering patterns that drive the observed phenomena. This information is valuable for producing refined analysis metrics for our implemented model for further analysis based on the identified causal relationships.

In our prediction algorithm, we assess the significance of causal links between pairs of features based on their p-values obtained from the Granger causality test. The p-value indicates the probability that the observed causal relationship could have occurred by random chance. To determine whether a causal link is considered significant, we compare the p-value to a predefined threshold. If the p-value for a pair of features is below this threshold, it suggests that the relationship is unlikely to be due to random fluctuations and thus can be deemed statistically significant.

By incorporating this criterion into our algorithm, we ensure that only robust and meaningful causal links are considered in our analysis. This approach allows us to focus on those relationships that demonstrate strong evidence of directional influence, enhancing the reliability of the predictions and interpretations of the proposed approach. Furthermore, by filtering out less significant links, we streamline our model to emphasize the most impactful feature interactions, leading to more accurate and precise predictions.

The LAMB (Layer-wise Adaptive Moments) optimizer, a variant of the Adam optimizer, is tailored for large-batch training by individually adjusting the learning rate for each layer based on the gradients and parameters, thereby ensuring optimised updates and mitigating instability. This feature is particularly beneficial for deep architectures like ResNet50, where varying sensitivities across layers necessitate distinct learning rates.

An adaptive learning rate scheduler OneCycleLR was incorporated to regulate the learning rate throughout training. This scheduler initiates with a higher learning rate, gradually reducing it as training progresses, which facilitates rapid convergence in the early phases and fine-tunes the model in the later stages, preventing overshooting of the optimal minima.

The experiments were conducted on a computational system equipped with an Intel Core i7 processor. The LAMB optimizer was employed for large-batch training, with a learning rate set to 0.001. An optimizer was incorporated ($\beta 1 = 0.9$, $\beta 1 = 0.9$, $\beta 2 = 0.98$, $\beta 2 = 0.98$) to ensure stable updates and prevent instability commonly encountered in deep architectures.

In addition to the LAMB optimizer, an adaptive learning rate scheduler, OneCycleLR, was implemented to further enhance the training process. The OneCycleLR scheduler was configured with a maximum learning rate of 0.01, a minimum learning rate of 0.0001, and a momentum value of 0.95, facilitating dynamic adjustment of the learning rate throughout the training procedure. Regularisation techniques, including a dropout rate of 0.5 and weight decay.

The combined application of the LAMB optimizer, learning rate scheduler, and regularisation techniques (including dropout and weight decay) led to notable improvements in model performance. The LAMB optimizer enhanced convergence speed and stability, the scheduler contributed to better generalisation, and the regularisation methods effectively mitigated overfitting by promoting a more balanced reliance on the model's features.

The architecture of the proposed network comprises several essential components designed to efficiently extract features from the input data and learn the underlying patterns. The network begins with the input layer, followed by the feature extraction modules. After feature extraction, fully connected layers are typically employed to perform high-level reasoning based on the extracted features. To mitigate overfitting and enhance generalisation, regularisation modules are incorporated, followed by optimisation and learning rate control mechanisms, culminating in the output layer.

IV. Result and Discussion

A. Experiment Results

In this section, we discuss the results of our work, presenting a comprehensive analysis through various figures and detailed interpretations. The figures illustrate key findings and trends, providing a visual representation of the data. By interpreting these results, we aim to provide a clear understanding of their implications and relevance to our research questions.

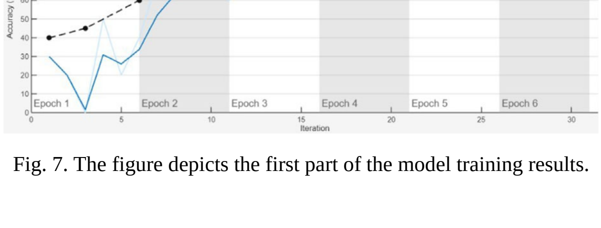

Fig. 7. The figure depicts the first part of the model training results.

Fig. 7 depicts a plot of the training and validation progress of the model. The y-axis represents the accuracy percentage, while the x-axis denotes the iterations across 6 epochs. The blue line tracks the training accuracy, while the black line represents the validation accuracy. The plot indicates that the model is learning effectively without significant overfitting, as the validation accuracy closely follows the training accuracy, especially in later stages. Fig. 8 illustrates the training and validation loss of the model. The x-axis represents the iterations, segmented into six epochs, while the y-axis shows the loss values. The orange line tracks the training loss, and the dashed black line represents the validation loss. This result demonstrates a successful training process, suggesting effective learning and minimal over-fitting.

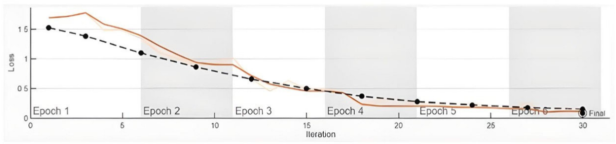

Fig. 8. The figure depicts the second part of the model training results.

The causal links visualization in Fig. 9 illustrates the relationships between different features extracted through the implemented model. Nodes in the figure represent individual features or components extracted from the image dataset. Edges represent causal relationships between these features. Each edge is an notated with a p-value derived from the Granger causality test, where a lower p-value (commonly less than 0.05) signifies a statistically significant causal relationship between the connected features. The presence of numerous causal links suggests that the model has identified a complex network of inter-dependencies between different image features. This indicates that the model leverages a rich and interconnected feature space, enhancing its ability to make accurate predictions.

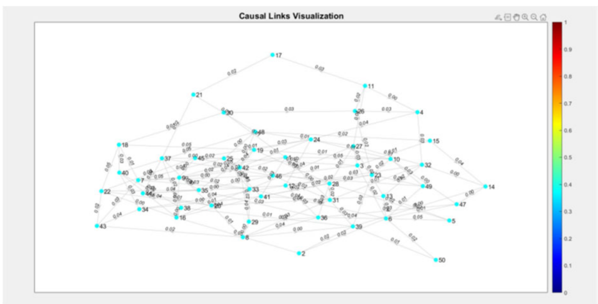

Fig. 9. Causal link graph based on the ImageNet dataset.

The 3D surface plot of causal coefficients offers a visual representation of the relationships between the features extracted from the dataset processed by the model shown in Fig. 9. This plot stems from a Granger causality analysis that was performed on the reduced feature set, extracted and processed through a pre-trained Resnet50 model and subsequently dimensionality reduced using Principal Component Analysis (PCA). The plot itself is a graphical depiction of how each feature potentially influences the other, with the surface colour and height indicating the magnitude and direction of these causal relationships.

Fig. 10 presents a 3D bar plot of a confusion matrix, visually depicting the performance of a localization model. Each bar represents the number of instances predicted for a particular class versus the actual class, with the "Predicted Class" and "Actual Class" axes indicating where predictions match or differ from true labels. The bar heights and color gradients (ranging from blue for lower counts to red for higher counts) represent the number of samples. Diagonal elements, such as TP11, TP22, and TP33, indicate true positive rates for each class, with taller, red-colored bars signifying accurate predictions. This 3D visualization highlights the model's performance across all classes.

B. Discussion

The variety and distribution of causal connections show that the model effectively captures and utilizes multiple aspects of the images, which are important for making accurate localizations or predictions. This robustness in feature extraction is crucial for high model performance. The features obtained were reduced using PCA, which retains the most significant features. The causal links between these principal components suggest that the model's core predictive capabilities are based on meaningful and causally related features.

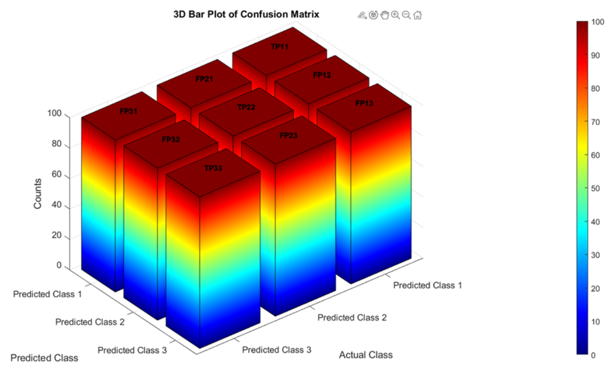

Fig. 10. Predicted results of the proposed model

The causal links identified have clearly had positive implications for the image prediction task, as shown in the result plots for the implemented model using Resnet50. Here are detailed insights into the algorithm:

- Enhanced Prediction Accuracy: By understanding the causal relationships between features, the model can better distinguish between different classes or categories in the image dataset, leading to improved prediction accuracy.
- Model Generalization: A model that understands the underlying causal structures of the data is likely to generalize better to new, unseen data. This means the model can maintain high performance even when applied to images that were not part of the training set.
- Interpretability: The visualization of causal links also aids in interpreting the model's decision-making process, which is valuable for debugging and refining the model.

The visualization of causal links in the implemented model offers a window into the intricate web of relationships between various image features. These links illustrate how the model perceives and processes the information contained within the image data to make accurate predictions. By mapping out these relationships, the visualization reveals the model's capability to identify which features are most influential in its decision-making process, allowing for a deeper understanding of the model's functioning. The presence of a dense network of significant causal links within the visualization is a testament to the effectiveness of our work.

This complexity in the model's architecture shows that it has been finely tuned to extract and use a wide array of features from the image data. Such a detailed network suggests that the model does not rely on a few prominent features but instead draws from a comprehensive and interconnected set of features, which enhances its ability to accurately classify or predict outcomes based on the input images. Overall, the causal links visualization not only highlights the sophistication of the proposed working mechanism but further showcases its high performance in image localization or prediction tasks. By effectively leveraging a rich feature space, the model demonstrates a robust capability to interpret and make sense of complex image data.

The predicted visualization of causal links serves as a valuable tool for model evaluation, offering insights that can guide further refinement and improvement of the model's predictive abilities.

The plot illustrates a varied distribution of causal coefficients, with peaks and valleys highlighting strong and weak causal relationships between features. Red peaks represent highly influential features critical to the model's decision making, while blue areas indicate features with minimal impact. This variability reflects a complex feature space where multiple features interact to enhance the model's predictive accuracy.

In the context of image prediction, the plot demonstrates that the model effectively leverages diverse features, rather than relying on a single one, showcasing its robustness and capability in handling advanced image localization tasks. The performance of the implementation indicates that the model generally performs well, particularly in correctly classifying instances of the target classes, as shown by the prominent bars along the diagonal. The plot suggests that while the model achieves high accuracy in predicting the majority of cases correctly, there are specific classes where the predictions could be more reliable.

The evaluation was conducted across multiple datasets to assess the performance of the proposed models as mentioned in Gayathri *et al.* [3] [7]. On the iNaturalist dataset, ResNet-101 achieved an accuracy of 85.40%, showcasing its ability to handle complex and large-scale fine-grained classification tasks effectively. For the PASCAL-50S dataset, which focuses on semantic segmentation and object recognition, ResNet-101 demonstrated a slightly higher accuracy of 86.20%, reflecting its adaptability to diverse image contexts. Using the Tiny ImageNet dataset, ResNet-50 attained an accuracy of 77.50%, indicating strong performance on smaller-scale datasets with challenging classification tasks. Lastly, on the EuroSAT dataset, which involves satellite image classification, ResNet-50 achieved an impressive accuracy of 98.60%, underscoring its robustness in high-dimensional spatial data. We have been able to demonstrate an accuracy of 99.7% for the Image Net dataset. These results collectively highlight the efficacy of our proposed architecture in achieving higher performance across varied datasets and domains.

V. Conclusion

We propose a comprehensive workflow for an image localization task, leveraging a pre-trained Resnet50 network as the backbone. The network is adapted to the specific task of enhanced prediction capabilities by modifying its architecture to incorporate advanced prediction mechanisms. The addition of causality can improve model performance by enabling more accurate predictions through understanding and leveraging cause-effect relationships. We have used this to improve the efficiency of the image localization tasks carried out using the ImageNet dataset. The implemented principal component analysis (PCA) algorithm retained the most significant features. These features are analyzed using Granger causality tests to explore potential causal relationships among the features. These advanced tests have been used as a mechanism to determine whether one variable can predict future values of another variable based on relevant information. Integrating a detailed feature analysis into the localization task leverages a novel workflow that enhances prediction by allowing the model to account for underlying dependencies in the data. We have developed our algorithm and tested the system accuracy on the ImageNet dataset. The methodological execution of algorithms ultimately leads to a more refined working mechanism for the model. This improves the model's overall predictive performance, making it more robust and interpretable in complex localization tasks.

Conflict of Interest

The author declares no conflict of interest.

References

[1] K. Balasamy and V. Seethalakshmi, "HCO-RLF: Hybrid classification optimization using recurrent learning and fuzzy for COVID-19 detection on CT images," *Biomedical Signal Processing and Control*, vol. 100, Part A, 106951, 2025. https://doi.org/10.1016/j.bspc.2024.106951

[2] K. Balasamy, V. Seethalakshmi, and S. Suganyadevi, "Medical image analysis through deep learning techniques: A comprehensive survey," *Wireless Pers. Commun.*, vol. 137, pp. 1685–1714, 2024. https://doi.org/10.1007/s11277-024-11428-1

[3] S. Deepa, J. L. Zeema, and S. Gokila, "Exploratory architectures analysis of various pre-trained image classification models for Deep learning," *Journal of Advances in Information Technology*, vol. 15, no. 1, pp. 66–78, 2024.doi: 10.12720/jait.15.1.66-78

[4] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. 2016 IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778.

[5] P. Gavrikov and J. Keuper, "Can biases in ImageNet models explain generalization," in *Proc. 2024 IEEE/CVF Conf. on Computer*

Vision and Pattern Recognition (CVPR), pp. 22184–22194. https://api.semanticscholar.org/CorpusID:268856816

[6] C. W. J. Granger. Investigating causal relations by econometric models and cross-spectral methods. [Online]. 1969. Available: https://api.semanticscholar.org/CorpusID:41012146

[7] T. Gayathri and K. Sundeep Kumar, "A deep learning based effective model for brain tumor segmentation and classification using MRI images," *Journal of Advances in Information Technology*, vol. 14, no. 6, pp. 1280–1288, 2023. doi: 10.12720/jait.14.6.1280-1288

[8] C. Sub-r-pa and R. Chen, "Knowledge distillation generative adversarial network for image-to-image translation," *Journal of Advances in Information Technology*, vol. 15, no. 8, pp. 896–902, 2024. doi: 10.12720/jait.15.8.896-902

[9] R. Lin, C. Dong, P. Zhou, P. Ma, S. Ma, X. Chen, and H. Liu, "Motor imagery EEG task recognition using a nonlinear Granger causality feature extraction and an improved slap swarm feature selection," *Biomed. Signal Process. Control*, vol. 88, Part A, 105626, 2024. https://doi.org/10.1016/j.bspc.2023.105626

[10] N. Mankovich, G. Camps-Valls, and T. Birdal, "Fun with flags: Robust principal directions via flag manifolds," in *Proc. 2024 IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2024, pp. 330–340.

[11] O. Russakovsky, J. Deng, H. Su *et al.*, "ImageNet large scale visual recognition challenge," *Int. J. Comput. Vis.*, vol. 115, pp. 211–252, 2015.

[12] MATLAB. [Online]. Available: https://in.mathworks.com/