Image and Vision Computing 148 (2024) 105147

Contents lists available at ScienceDirect

Image and Vision Computing

journal homepage: www.elsevier.com/locate/imavis

Two-dimensional hybrid incremental learning (2DHIL) framework for semantic segmentation of skin tissues

Muhammad Imran [a], Muhammad Usman Akram [b,*], Mohsin Islam Tiwana [a], Anum Abdul Salam [b], Taimur Hassan [c], Danilo Greco [d]

[a] Department of Mechatronics Engineering, National University of Sciences and Technology, Islamabad, Pakistan
[b] Department of Computer and Software Engineering, National University of Sciences and Technology, Islamabad, Pakistan
[c] Dept of Electrical and Computer Engineering, Abu Dhabi University, Abu Dhabi, United Arab Emirates
[d] Department of Management, Economics and Industrial Engineering, Politecnico di Milano, Via Lambruschini 24/b, 20156 Milan, Italy

ARTICLE INFO

Keywords:
AI
Incremental semantic segmentation
Incremental learning
Continual learning
Knowledge distillation
Mutual distillation loss
Transformer
Segformer
Skin cancer
Non-melanoma skin cancer
Computational histopathology

ABSTRACT

This study aims to enhance the robustness and generalization capability of a deep learning transformer model used for segmenting skin carcinomas and tissues through the introduction of incremental learning. Deep learning AI models demonstrate their claimed performance only for tasks and data types for which they are specifically trained. Their performance is severely challenged for the test cases which are not similar to training data thus questioning their robustness and ability to generalize. Moreover, these models require an enormous amount of annotated data for training to achieve desired performance. The availability of large annotated data, particularly for medical applications, is itself a challenge. Despite efforts to alleviate this limitation through techniques like data augmentation, transfer learning, and few-shot training, the challenge persists. To address this, we propose refining the models incrementally as new classes are discovered and more data becomes available, emulating the human learning process. However, deep learning models face the challenge of catastrophic forgetting during incremental training. Therefore, we introduce a two-dimensional hybrid incremental learning framework for segmenting non-melanoma skin cancers and tissues from histopathology images. Our approach involves progressively adding new classes and introducing data of varying specifications to introduce adaptability in the models. We also employ a combination of loss functions to facilitate new learning and mitigate catastrophic forgetting. Our extended experiments demonstrate significant improvements, with an F1 score reaching 91.78, mIoU of 93.00, and an average accuracy of 95%. These findings highlight the effectiveness of our incremental learning strategy in enhancing the robustness and generalization of deep learning segmentation models while mitigating catastrophic forgetting.

1. Introduction

1.1. AI models

AI has gone miles and manifested its impact in almost every field of life since its introduction. It has made life easier in day-to-day living and enhanced efficiency in professions, businesses, and academia. Its impact is particularly notable in healthcare and diagnostics. While AI with its deep learning models has been able to match human performance in certain cases and even surpassed in a few, there is a visible limitation of these models when compared with humans; these are incapable of evolution. While humans continuously grow with time in their abilities and knowledge, AI models fail to evolve. Hence, their application is limited not only to a particular set of tasks but also to types of test cases. AI models perform only those tasks they have been trained for and perform well on test cases similar to their training data.

1.2. Growth and robustness of AI models

The ability to grow in capabilities and perform more tasks with acceptable performance is desired from AI models irrespective of their field of application. AI, whether it is employed for self-driving cars,

DOIs of original article: https://doi.org/10.1016/j.imavis.2024.105148, https://doi.org/10.1016/j.imavis.2024.105098.
* Corresponding author.
E-mail address: usmakram@gmail.com (M.U. Akram).

https://doi.org/10.1016/j.imavis.2024.105147

Available online 3 July 2024

virtual assistance, business analytics, health care, or digital diagnostics, faces new challenges every day. The discovery of new challenges and anomalies leaves well-trained and previously well-capable AI models redundant due to their inability to evolve. Many efforts have been made to enable models to evolve through incremental learning. However, all of these efforts are seriously challenged by their inability to retain previously learned knowledge while learning new one. These models tend to forget previously gained knowledge while learning new and the attribute is called catastrophic forgetting [1,2]. Noticeably, while most initial efforts for incremental learning focused on classification models [3–10], segmentation models have also been experimented with incremental learning to enable them to segment new types [11–15].

2. Related work

2.1. Incremental learning

AI models experience catastrophic forgetting when trained for new tasks. To address this challenge, several incremental learning approaches have been developed. One commonly adopted approach is knowledge distillation [16], where a new model is trained to replicate the outputs or representations of the original model while learning new knowledge [4,14,17–20]. Another approach involves storing exemplars from representative classes [5,15,21–24], but the selection of appropriate exemplars is crucial. Zhuang et al. [25] suggest criteria for selecting suitable exemplar sets for replay. Smith et al. [26] and Maracani et al. [27] suggest the use of synthetic data for replay instead of storage. Dynamically expandable networks have also been proposed for incremental learning [28–30]; these suggest adding a new component at each subsequent stage of learning along with a method to scale the network. Joint training of new models using both old and new data to overcome catastrophic learning has also been used [31]. This is an exhaustive approach; however, if the major portion of training data belongs to new tasks with only a small proportion belonging to old ones, then it really helps alleviate catastrophic forgetting. The classes, however, that have become obsolete and redundant should be omitted from training data. Besides catastrophic forgetting and unlike classification, incremental semantic segmentation faces a critical issue of background shifting. Background shifting is a phenomenon where a segment identified by the original segmentation model as background may belong to a new class introduced later through the incrementally trained model. Pseudo-labeling is generally the most common approach to address background shifting where background regions of new training images are labeled based on the primary knowledge of the original model [13]. Another approach is to identify feature-rich salient segments and identify them as unknown classes rather than the background [15]. Incremental learning can be classified into three scenarios: task-incremental learning, domain-incremental learning, and class-incremental learning. Task incremental learning is a learning paradigm where a model sequentially learns to solve a series of distinct tasks, with each task presenting a unique learning objective or problem to be solved. This approach enables the model to gradually accumulate knowledge and adapt its capabilities over time to perform new tasks. Domain incremental learning focuses on training a model to solve problems in varying contexts and domains. In this paradigm, the model learns to generalize across different environments or datasets, allowing it to perform effectively in diverse real-world situations. In class incremental learning, the model incrementally learns to discriminate among newly observed classes while retaining knowledge of previously learned classes. This learning strategy enables the model to adapt to an expanding set of classes over time, facilitating continuous improvement and refinement of its classification abilities [32]. Our proposed work belongs to the class-incremental learning scenario, believed to be the most challenging for deep neural networks, particularly without having a memory component.

2.2. Semantic segmentation

Automatic semantic segmentation is a computer vision task requiring each test image pixel to be identified as one of the differentiable categories, including the background. However, semantic segmentation cannot determine whether two different pixels belonging to a similar category are part of the same object. It is a core task in many computer vision applications such as self-driving cars, autonomous robots, surveillance, medical diagnostics, and image analysis. Many notable semantic segmentation models have been introduced over time and among these, fully convolutional network (FCN) [33] pioneers the modern semantic segmentation approaches. It employs convolutional layers instead of fully connected layers, incorporates skip connections for feature extraction, and has demonstrated high performance at semantic segmentation tasks. U-net [34] is well known for its ability to identify subtle and intricate details in images, making it suitable for processing high-resolution images and thus has been widely used in biomedical applications. A few more notable models include Deeplab series [35–37], Segnet [38], Efficient Neural Network (ENet) [39], ResNet [40], Pyramid Scene Parsing Network (PSPNet) [41] and EfficientNet [42].

Since their introduction, transformer-based segmentation models and approaches utilizing attention mechanisms have outperformed conventional CNN models. Vision Transformer (ViT) [43], Segformer [44], Swin Transformer [45], SETR [46], and Maskformer [47] have shown promising results at semantic segmentation. The ability of transformer-based models to capture global context, long-range dependencies, and relationships between pixels in an image proves them well-suited for most segmentation tasks.

In addition to the previously highlighted models, TransUnet [48] is another prominent hybrid model that combines the strengths of both transformer and U-Net. The high-resolution feature maps of U-Net, combined with the global context extracted by the transformer, contribute to its promising performance at segmentation tasks. Pyramid Vision Transformer (PVT) [49] capitalizes on the best of both CNN and transformer architectures, making it well-suited for high-resolution tasks with relatively low computational requirements.

2.3. Incremental semantic segmentation

The segmentation models discussed earlier require knowledge of classes and training data during the training process, which is conducted in a single iteration. This constraint limits their robustness and ability to effectively adapt to new tasks, despite their demonstrated performance at previous tasks. An attempt to update these models for new classes is challenged by catastrophic forgetting, which imposes serious restrictions on maintaining prior performance. Additionally, these models exhibit high sensitivity to variations between test data specifications and the training data. Given the dynamic nature of real-world applications, characterized by the continual emergence of new classes and diverse data, there is a growing demand for incremental learning approaches. Incremental learning aims to develop segmentation models capable of adapting to new classes, tasks, and domains, thereby enhancing robustness across different test data specifications.

As already highlighted, most incremental learning research has focused on classification tasks [3–8,21]. However, recognizing the limitations of existing semantic segmentation algorithms and the potential benefits of incremental learning, there has been increasing research interest in incremental semantic segmentation.

Incremental semantic segmentation not only faces the challenge of catastrophic forgetting like incremental classification but also encounters the additional challenge of background shifting. Among many approaches proposed so far, knowledge distillation is the most common approach to address catastrophic forgetting in which a new model is made to replicate the results, features, or representations of the old model. Kirkpatrick et al. [9] employ two loss functions for knowledge

distillation, one for replicating the intermediate feature space and another for the output space. Michieli et al. [12] enforce knowledge distillation by replicating the latent space of the new model with the older one. Besides, it sparsifies features to accommodate the learning of new classes and employs contrastive learning for class clustering. Zhao et al. [50] propose knowledge distillation at the intermediate and output layers along with contrastive learning to address semantic distribution shift. Michieli et al. in their work [19,51] have proposed output and feature-level distillation losses for knowledge distillation. Additionally, they recommend freezing certain parts of the model to enhance stability by preventing learning. Douillard et al. [13] introduce a distillation technique that matches long-term and short-term dependencies of models at the feature level by comparing output at each intermediate layer. Oh et al. [52] suggest an adaptive logit regularizer that alleviates catastrophic forgetting while maximizing the learning of new classes. Kalb et al. [53] identify biased learning of new classes and semantic shifts in the background as the main causes of catastrophic forgetting, proposing an unbiased cross-entropy loss for classification and knowledge distillation as the solution. Cermelli et al. [14] also highlight the issue of semantic distribution shift associated with incremental semantic segmentation and propose a distillation framework to tackle it. Moreover, it suggests a method to initialize classifier parameters to ensure unbiased predictions for both background and new classes. Cha et al. [15] prevent forgetting by freezing part of the previous model and employ pseudo-labeling to mitigate catastrophic forgetting. Zhang et al. [54] suggest knowledge distillation using a representation compensation module with two branches, one frozen for stability and the other trainable for learning new knowledge. The method additionally, employs 3D knowledge distillation by pooling spatial and channel information.

Another approach to preserve the knowledge of previously trained models involves replaying examples belonging to old classes. Researchers also employ this technique to preserve knowledge in semantic segmentation tasks. Cha et al. [15] integrate the replay strategy with knowledge distillation by storing a selected, class-balanced set of past data as exemplars. In addition to knowledge distillation and parameter freezing, Kirkpatrick et al. [9] also suggest replaying a small portion of data belonging to old tasks to preserve knowledge. Oh et al. [52] suggest extracting features for each class, averaging them on a per-class basis, and storing a specific number of features for replay during training at each incremental stage. Yu et al. [55] also suggest rehearsals for retaining knowledge about old classes. Unlabeled data is pseudo-labeled by both the old model and a temporary model trained for new tasks and fusion of these pseudo-labeled data is employed for replay. Maracani et al. [27] collect replay data for past classes through web sources and also generate it through GAN. This replay data is then combined with new data, and the new model is jointly trained using both old and new classes. As already highlighted background shift is one of the major challenges in class incremental semantic segmentation [53]. Goswami et al. [56] suggest the selection of the most relevant channels from the background of the previous model and transferring only these weights to the new model. The approach addresses the background shift problem and enhances the plasticity of the new model. This improvement aids in predicting new classes, mitigates background semantic shift, and alleviates catastrophic forgetting. Cermelli et al. [57] explicitly suggest methods to address the background shift problem by introducing new classification and distillation losses. These losses incorporate both the new classes and previously seen classes that appeared as background in new training samples. Additionally, it tackles biased prediction by introducing a novel weight initialization method at each incremental step. Yang et al. [58], Douillard et al. [13], and Cha et al. [15] address background shift through pseudo-labeling.

After the success of transformer models in semantic segmentation, they are now being utilized in incremental semantic segmentation settings. A pioneering attempt in this direction is [59], which employs transformers for incremental semantic segmentation. It exploits the ability of Segformer [44] to generate self-attention maps and utilizes them for knowledge transfer to new models. Similarly, Shang et al. [60] employed Segmenter [61] where the transformer decoder only requires new class tokens for learning. Besides, the approach suggests an unbiased classification strategy and a distillation approach that solely focuses on the old classes, thereby improving stability. Apart from experimental settings, incremental semantic segmentation has also been employed in practical scenarios like cardiac image segmentation [62] and diabetic retinopathy lesion segmentation [63].

Skin cancer is a significant global health concern with high incidence rates worldwide. In cases where it is diagnosed early, skin cancers often demonstrate high rates of recovery. However, the contrast between lesion and non-lesion areas in skin cancer images tends to be low. Therefore, distinguishing between these areas requires expertise, often leading to differences of opinion among dermatologists during diagnosis. Therefore, computer-aided automatic diagnostic systems are critical for successfully diagnosing skin lesions. However, the automatic analysis of human skin is challenging due to variations in features such as roughness, tone, and hair mass, which depend on various factors [64]. Automatic skin lesion analysis typically involves four stages: preprocessing, segmentation, feature extraction, and classification [65]. The effectiveness of classification relies heavily on the outcomes of the preceding stages [66]. Segmentation of skin lesions is one of the critical methods for developing an accurate diagnostic system. Segmentation is the process of separating a lesion area from a non-lesion area in an image [67]. Successful segmentation enhances the contextual understanding of features and diseases needed for the subsequent classification step [68]. Research by Thomas et al. [69], Öztürk et al. [70], Wu et al. [71], Kleczek et al. [72], and another study by Wu et al. [73] has lately contributed to skin lesion segmentation, yielding promising results. However, their results still fall short of expectations.

In this work, we introduce a 2-dimensional hybrid incremental learning (2DHIL) framework for semantic segmentation. It combines knowledge distillation and joint training to overcome catastrophic forgetting. The capability of the proposed model grows incrementally in two dimensions. It not only learns to segment more classes but also grows to test heterogeneous samples of different specifications. We propose a combination of loss functions for new learning and knowledge distillation. The proposed knowledge distillation approach employs self-attention feature maps extracted through multiple transformer blocks of the hierarchical encoder proposed in Segformer [44]. Moreover, joint training has also been incorporated for incremental training of the new model with a major portion of training samples from new classes and a small portion from old classes. It is however highlighted that data belonging to old classes has not been previously used in the training of the original model. The proposed model has been tested through a complex practical problem of segmenting non-melanoma skin cancer and tissues from histopathology images. The results of our tests advocate that the method can be successfully employed for semantic segmentation in practical applications. The method is particularly useful for applications where the problem set is expected to change by the discoveries of new challenges or in cases where specifications of test samples may vary due to progress in imaging techniques. The contributions of the study are listed as follows:

- We propose a novel 2-dimensional hybrid incremental learning framework. This framework employs an incremental semantic segmentation loss, developed by combining classification loss, mutual distillation loss, and KL divergence. The incremental semantic segmentation loss ensures new learning and knowledge distillation, effectively mitigating catastrophic forgetting. Additionally, the framework employs joint training with both previously known and new classes, in appropriate proportions.
- To the best of our knowledge, this is the first incremental learning semantic segmentation model that employs Segformer for a practical

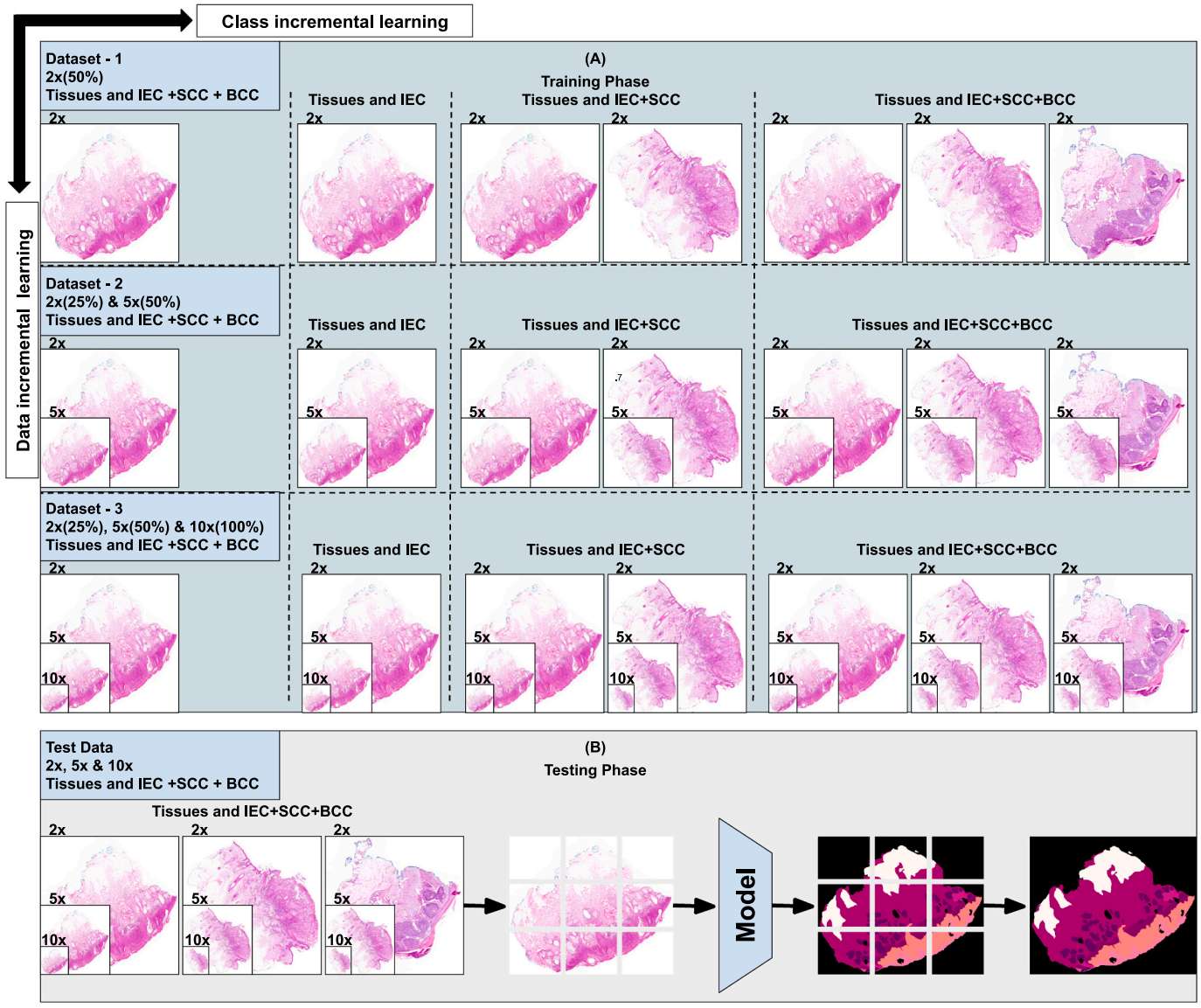

Fig. 1. 2-Dimensional hybrid incremental learning (2DHIL) framework for incremental semantic segmentation. (A) Schematic explanation of the training phase where class incremental learning and data incremental learning progress along the horizontal axis and vertical axis respectively. (B) The testing phase of the model where the test set includes data of all the classes and specifications.

application and demonstrates performance by segmenting non-melanoma skin cancer and tissues.

- This study employs a mutual distillation loss function, for the first time, in an incremental semantic segmentation problem.

3. Methodology

Our proposed incremental learning framework for semantic segmentation, named 2-dimensional hybrid incremental learning (2DHIL), is configured to train AI models simultaneously in two directions as shown in Fig. 1. The learning along the horizontal axis is class incremental learning and along the vertical axis, it is data incremental learning.

3.1. Class incremental learning

Our framework is trained to gain more knowledge and identify more classes along the horizontal axis as depicted in Fig. 1. This learning is class incremental learning enhancing the capability of the model and enabling it to segment more classes and anomalies. The model is originally trained to identify skin tissues and Intraepidermal carcinoma (IEC) with training samples and masks of a particular resolution. From this point onward training proceeds in two directions such that along the horizontal axis, the model is trained to identify new classes. Moving along the horizontal axis in the next step, the model is incrementally trained to recognize a new class of squamous cell carcinoma (SCC) in addition to already learned classes. In this phase of training the model is presented with training data and segmentation masks of both previous and new classes but the major portion of training data belongs to the new class. In the subsequent step, the model is again incrementally trained in the same direction for the identification of a novel class, basal cell carcinoma (BCC) with training data of BCC in greater proportion however a smaller portion of data pertaining to old classes is also used to overcome catastrophic forgetting through replay.

3.2. Data incremental learning

In data incremental learning, the model is trained to process and

Training Strategy									
Data Resolution	Phase-I			Phase-II			Phase-III		
2x	50%			25%			25%		
5x	-			50%			50%		
10x	-			-			100%		
Classes	Step-I	Step-II	Step-III	Step-I	Step-II	Step-III	Step-I	Step-II	Step-III
Skin Tissues & IEC	50%	25%	25%						
SCC	-	50%	50%						
BCC	-	-	100%						
Skin Tissues & IEC				50%	25%	25%			
SCC				-	50%	50%			
BCC				-	-	100%			
Skin Tissues & IEC							50%	25%	25%
SCC							-	50%	50%
BCC							-	-	100%

Fig. 2. Proposed composition of new and old data for joint training at each incremental stage of learning.

identify classes from data of varying specifications. This learning makes the model robust addressing its sensitivity to data specifications. Fig. 1 shows data incremental learning along the vertical axis. Data incremental learning for our proposed model is similar to class incremental learning. Firstly, the model is trained to identify classes through data of a particular resolution. Later, in each subsequent step training data of both new and old resolutions is used with a major portion of the new resolution thus enabling the model to handle data of varying specifications.

Moreover, it is important to notice that training in a particular direction that is toward class incremental learning or data incremental learning is not strictly necessary; it may take the course as need be depending on the discovery of a new class or data of a new specification.

3.3. Training phase

We incrementally train our model to learn new knowledge while preserving the previously gained one. We employ a hybrid approach to address catastrophic forgetting, the biggest challenge in incremental learning. We, for the first time, combine mutual distillation loss function in incremental semantic segmentation problem for knowledge distillation along with replay of old classes. We ensure the replay of previously trained classes at each incremental learning stage. Therefore, the training data at each incremental stage includes both new classes (not previously trained) and old classes in proportion, as depicted in Fig. 2. As shown in Fig. 2, the training strategy is implemented in three phases: Phase-I, Phase-II, and Phase-III. Each phase employs data from all classes but of different resolutions. These phases correspond to data incremental learning. Furthermore, each phase is subdivided into three steps, each corresponding to class incremental learning. As depicted in Fig. 2 Phase-I uses 50% training data of 2x resolution. The training data of this phase is further divided class-wise in each step. Step-I employs 50% of IEC training samples to train the model for the identification of skin tissues and IEC. It is worth noting that in the initial stage of training, all training data is entirely new, and the model has not been previously trained to identify any of the classes presented; this data is referred to as new data. In step-II, 25% of the data belonging to IEC of 2x resolution is used. As the model has already been trained to identify IEC from images of 2x resolution, this data is considered old data. Additionally, 50% of the data for SCC also forms part of the training data and is referred to as new data. Similarly, Fig. 2 illustrates the composition of new and old data for training at each stage of incremental learning.

3.3.1. Knowledge distillation and new learning

In each training phase, whether incrementing to learn new classes or adapting to data of new specifications, we utilize a combination of three loss functions to enforce three constraints: classification loss, mutual distillation loss, and Kullback–Leibler (KL) divergence. The incremental semantic segmentation loss L_{iss} is defined as follows:

$$L_{iss} = L_c + L_{md} + L_{KL} \tag{1}$$

L_c here is the classification loss, L_{md} is the mutual distillation loss and L_{KL} is the KL divergence.

3.3.2. Classification loss

The classification loss L_c is computed using categorical cross-entropy, ensuring that the model learns to identify and segment all classes presented in the current training phase. As our dataset includes both new and old classes, the model reinforces previously learned classes while learning new ones. The classification loss is defined as follows:

$$L_c = -\frac{1}{b_s}\sum_{i=0}^{b_s-1}\sum_{j=0}^{N_{cl}-1}\sum_{w=0}^{W-1}\sum_{h=0}^{H-1} Rep_{new}^{true}(x_{i,j,w,h})\log p(Rep_{old}|cl = cl_{i,j,w,h}) \tag{2}$$

here b_s represents the batch size and N_{cl} denotes the total number of classes in the current training phase. W and H stand for the width and height of the image, respectively. Rep_{new}^{true} is the representation of a label pertaining to a new class, Rep_{old} is the representation belonging to a previously learned class and cl represents classes. Moreover, the input at

each stage of incremental learning is a color image x such that $x \in \mathbb{R}^2$ that includes training data for both old and new classes and of all specifications.

3.3.3. Mutual distillation loss

The Mutual distillation loss L_{md}, as proposed by [18], employs Bayesian inference to evaluate structural and semantic similarities between representations of the original and incrementally learned knowledge. It jointly optimizes these representations, ensuring that while learning new knowledge, the original knowledge of the model is preserved.

L_{md} is calculated by jointly optimizing the distributions of previous and new representations as follows:

$$p(Rep_{old}, Rep_{new}) = p(Rep_{old}|Rep_{new}) \times p(Rep_{new}) \tag{3}$$

$$p(Rep_{new}, Rep_{old}) = p(Rep_{new}|Rep_{old}) \times p(Rep_{old}) \tag{4}$$

Where Rep_{new} and Rep_{old} are the feature representations belonging to new and old knowledge, respectively.

Incorporating the distributions pertaining to representations of knowledge in model training enables the model to understand the relationship between both newly acquired knowledge and that of the original model. This ensures that the model optimally learns new knowledge while preserving the one already gained by the original model. Since this knowledge and these representations are about different classes, incorporating classes in Eqs. (3) and (4) results as follows:

$$p(Rep_{old}, Rep_{new}|cl) = \sum_{i=0}^{N_{cl}-1} p(Rep_{old}|Rep_{new}, cl = cl_i) \times p(Rep_{new}|cl = cl_i) \tag{5}$$

$$p(Rep_{new}, Rep_{old}|cl) = \sum_{i=0}^{N_{cl}-1} p(Rep_{new}|Rep_{old}, cl = cl_i) \times p(Rep_{old}|cl = cl_i) \tag{6}$$

N_{cl}, as already indicated, is the number of classes.

The class conditional probabilities for representations given particular classes are approximated through multivariate Gaussian distribution as follows:

$$N\left(\hat{y}|\mu, \sum\right) = \frac{1}{\sqrt{2\pi^d|\sum|}} exp - \frac{1}{2}(\hat{y} - \mu)^T \sum^{-1} (\hat{y} - \mu) \tag{7}$$

where $\hat{y} \in y$ and $y = \{x|x \in \mathbb{R}^2\}$ are the color segmentation masks belonging to training samples of both old and new classes, d is the dimension of data, μ is mean and $\sum$ is the covariance of distribution and calculated as follows:

$$\mu = \frac{1}{N_t} \sum_{i=0}^{N_{t-1}} \hat{y} \tag{8}$$

and

$$\sum = \frac{1}{N_t - 1} \sum_{i=0}^{N_{t-1}} (\hat{y} - \mu)(\hat{y} - \mu)^T \tag{9}$$

Assuming N_{cl_i} represents the number of data samples belonging to class cl_i and N_t denotes the total number of data samples in the dataset belonging to all classes, then the prior probability of class cl_i, can be calculated as:

$$p(cl = cl_i) = \frac{N_{cl_i}}{N_t} \tag{10}$$

Now, Bayes' rule is used to calculate the posterior for each class cl_i, using class conditional probabilities computed using Eq. (7) and prior probabilities computed using Eq. (10).

$$p(cl = cl_i|Rep_{old}, Rep_{new}) = \frac{p(Rep_{old}, Rep_{new}|cl = cl_i)p(cl = cl_i)}{\sum_{i=0}^{N_{cl}-1} p(Rep_{old}, Rep_{new}|cl = cl_i)} \tag{11}$$

$$p(cl = cl_i|Rep_{new}, Rep_{old}) = \frac{p(Rep_{new}, Rep_{old}|cl = cl_i)p(cl = cl_i)}{\sum_{i=0}^{N_{cl}-1} p(Rep_{new}, Rep_{old}|cl = cl_i)} \tag{12}$$

Taking log on both sides of 11 results:

$$\log p(cl = cl_i|Rep_{old}, Rep_{new}) = \log \frac{(Rep_{old}, Rep_{new}|cl = cl_i) \times p(cl = cl_i)}{(Rep_{old}, Rep_{new}|cl = cl_i} \tag{13}$$

and L_{md} for class cl_i is calculated by jointly minimizing $\log p(cl = cl_i|Rep_{old}, Rep_{new})$ and $\log p(cl = cl_i|Rep_{new}, Rep_{old})$:

$$\begin{aligned} L_{md}(cl_i) = & -\sum_{j=0}^{n_o-1} Rep_{old}^{true}(x_j) \Big/ \tau \log p(cl \\ & = cl_i|Rep_{old}, Rep_{new}) - \sum_{j=0}^{n_n-1} Rep_{new}^{true}(x_j) \Big/ \tau \log p(cl \\ & = cl_i|Rep_{new}, Rep_{old}) \end{aligned} \tag{14}$$

where τ is the scaling factor, n_o and n_n are the number of old and new samples respectively.

3.3.4. KL divergence

Additionally, we employ KL divergence loss in our model to support new learning. This loss measures the disparity between the model's predictions and the true labels by comparing their probability distributions. Minimizing KL divergence aligns the model's predictions with the true labels. KL divergence is given as:

$$L_{KL} = \frac{1}{b_s} \sum_{i=0}^{b_s-1} \sum_{j=0}^{N_{cl}-1} \sum_{w=0}^{W-1} \sum_{h=0}^{H-1} Rep_{new}^{true}(x_{i,j,w,h}) \log \left(\frac{Rep_{new}^{true}(x_{i,j,w,h})}{p(Rep_{new}|cl = cl_{i,j,w,h})} \right) \tag{15}$$

3.3.5. Replay through joint training

To ensure the replay of learned classes through joint training, we propose that training data at each incremental step should include both old and new classes. Given that the old classes have already been learned and previous knowledge of the model is also preserved via knowledge distillation, we recommend that data from new classes should constitute the majority. It is also pertinent to mention that our approach does not require any memory component as we do not recommend storing any data or representations rather we propose using unseen data of old classes in a limited number for replay. We further propose to eliminate the replay data for the classes that become obsolete over time.

3.4. Testing phase

The final model, incrementally trained for both class and data incremental learning using our proposed framework, not only can segment more classes than the original model but it can also handle data of varying specifications. Our trained model can segment skin tissues and all three types of non-melanoma skin cancers i.e. IEC, SCC, and BCC without any strict constraint on the resolution or specifications of test data. Therefore, our proposed incremental learning framework makes a segmentation model robust, versatile and provides it with the capability to evolve with time. It is worth noting that during testing, we employ a single-head evaluation strategy, where test samples from all classes and data specifications are evaluated together and scored accordingly.

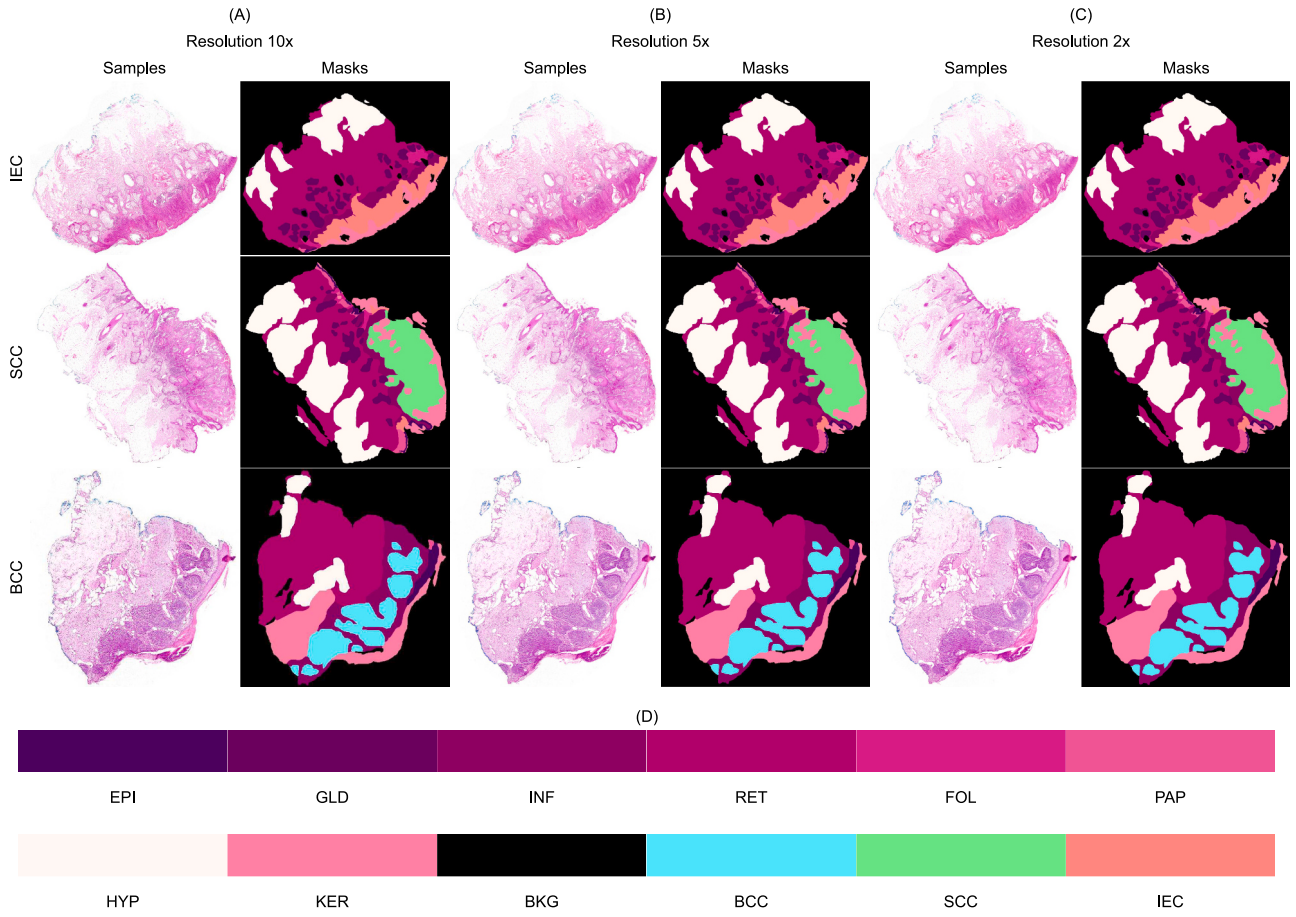

Fig. 3. Multi-resolution samples of skin tissues and non-melanoma skin cancers (IEC, SCC, and BCC), corresponding segmentation masks, and color palette. (A) Data samples and corresponding masks of 10x resolution. (B) Data samples and corresponding masks of 5x resolution. (C) Data samples and corresponding masks of 2x resolution. (D) All 12 classes with corresponding color palettes.

4. Experiment

4.1. Dataset

The dataset provided by [74], comprising 290 histopathology images along with corresponding manually created segmentation masks, has been used in our experiments. The dataset contains 290 slides indicating specific tissue sections that represent typical instances of non-melanoma skin cancer. In the dataset, there are 140 slides of BCC, 60 slides of SCC, and 90 slides of IEC, which represent three different types of non-melanoma skin cancers in different proportions. Also, the specimens have been obtained differently; 132 specimens have been extracted through excision biopsies, 100 through shaving and the remaining 58 have been obtained through punches. The sections of tissues that are most representative of cancer have been indicated on each slide by the pathologist. The data collection effort spans over four months starting from late 2017 to early 2018. The data has been collected from patients with ages spread over a wide range; the youngest being 34 and the oldest being 96 with a median age of 70. Besides, male to female ratio among patients is 2:1. The slides of biopsy samples have been imaged at high resolution, preprocessed, and saved as TIF images.

4.2. Segmentation and ground truth

The dataset has 12 distinguishable classification categories which include three types of carcinomas, multiple skin tissues, and background. BCC, SCC, and IEC are the three carcinomas and skin tissues include glands (GLD), hair follicles (FOL), inflammation (INF), reticular dermis (RET), hypodermis (HYP), papillary dermis (PAP), epidermis (EPI) and keratin (KER). Besides these, background (BKG) is also identified as a different category. A diligent effort of approximately 250 h has been put in by trained professionals and pathologists to create segmentation masks for these slides. While creating these masks, each pixel of the slide has been identified as one of the twelve categories and painted in a different color. Therefore, each pixel of the resultant mask has one of the twelve colors for which ImageJ software has been used. These segmentation masks have been saved in PNG format. All these categories and respective colors are shown in Fig. 3.

It is interesting to note that while identifying boundaries between tissues and carcinomas, pathologists do not work on a fixed magnification level rather magnification is varied depending on the complexity of the specimen and density of co-located categories. Moreover, different imaging equipment held at different labs does not necessarily have similar magnification capacity hence collection of data on a particular resolution is although possible but can be very difficult and impractical. Moreover, a model trained on a fixed resolution or magnification level may not be accurate enough when presented with a specimen of a different resolution. To make a model robust and accurate for specimens collected at different magnifications and resolution levels by different imaging equipment, it is important to train it on data of varying magnification and resolution. This is why our proposed framework makes use of down-sampled images of original data. The original data has been down-sampled by a factor of 2, 5, and 10 and named as 2x, 5x and 10x datasets.

A visual analysis of the same samples of 10x, 5x, and 2x as shown in Fig. 3 reveals that no visible difference is noticed in most cases however,

Table 1
Impact of variation in specifications of training data and test data on performance of segmentation models.

Test Data Resolution	Model	Training Data 10x			Training Data 5x			Training Data 2x		
		F1 Score	mIoU	Average Accuracy	F1 Score	mIoU	Average Accuracy	F1 Score	mIoU	Average Accuracy
2x	Segformer	88.34	88	86.45	86.38	85	88.9	**89.78**	**86**	**90.98**
5x	Segformer	88.79	89	89.78	**89.56**	**88**	**91.05**	86.78	83	85.78
10x	Segformer	**90.89**	**89**	**92.00**	88.76	88	89.09	88.76	88	87.78

extremely minute details may be omitted in certain samples of the 10x dataset.

4.3. Training and implementation details

We implemented Segformer [44] as the backbone for our incremental semantic segmentation framework using Python 3.11.4. The incremental semantic segmentation loss L_{iss}, a combination of loss functions as described in Eq. (1), has been employed for training the model. L_{iss} is developed by combining the classification loss, L_c, mutual distillation loss, L_{md}, and KL divergence L_{KL}, as described in Eqs. (2), (14) and (15) respectively.

We trained our model using NVIDIA GeForce RTX 2070 of 8GB. The training of the model has been organized in a single batch with 50

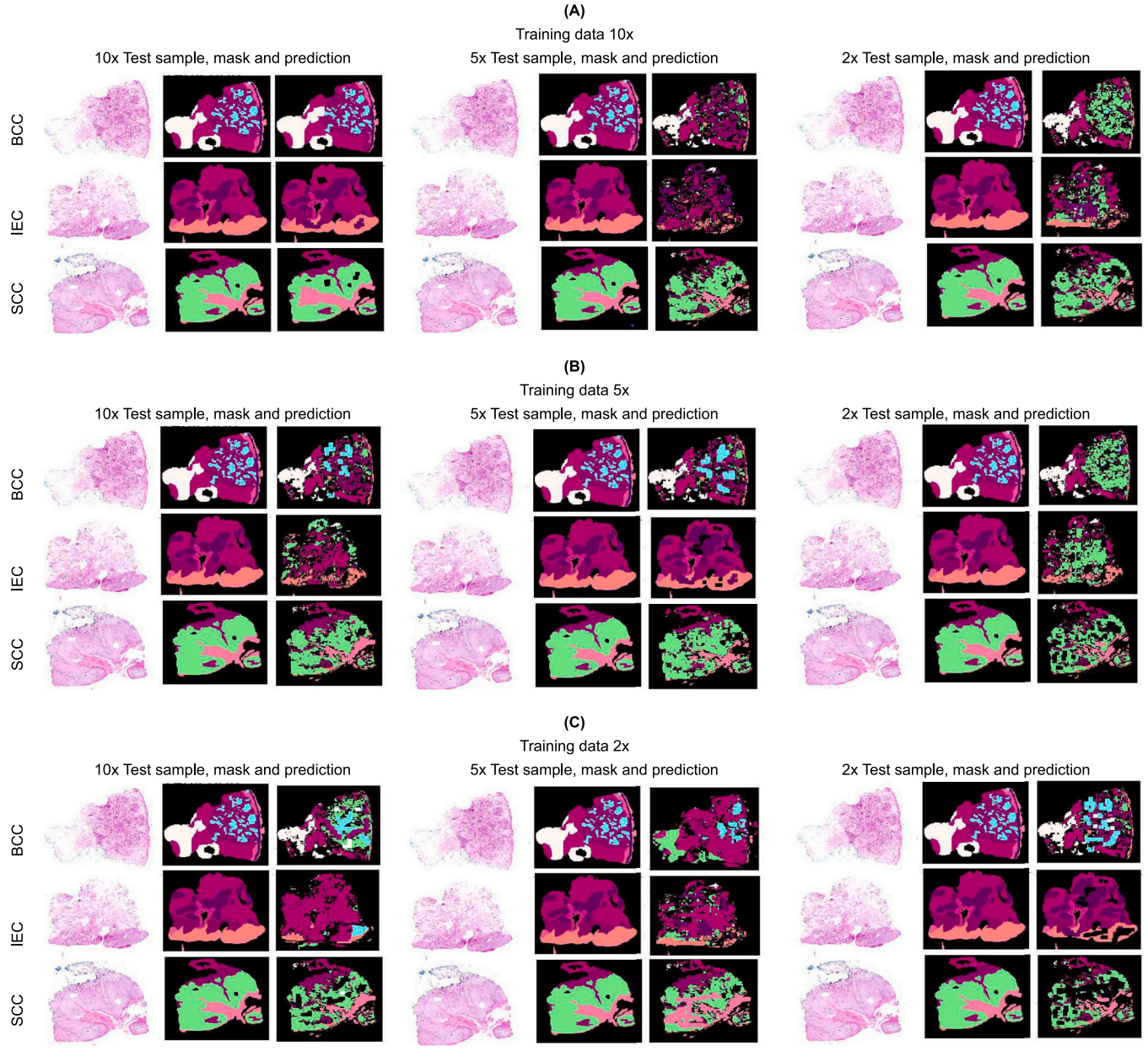

Fig. 4. Multi-resolution samples of non-melanoma skin cancers (BCC, IEC, and SCC), corresponding segmentation masks and predictions by the model. (A) Predictions for test cases of all cancers and all resolutions by the model trained on 10x resolution. (B) Predictions for test cases of all cancers and all resolutions by the model trained on 5x resolution. (C) Predictions for test cases of all cancers and all resolutions by the model trained on 2x resolution.

Table 2
Ablation studies: 2DHIL(Partial) - Training partial layers and 2DHIL - training all layers.

Test Data Resolution	Model	Training Data								
		10x			5x			2x		
		F1 Score	mIoU	Average Accuracy	F1 Score	mIoU	Average Accuracy	F1 Score	mIoU	Average Accuracy
2x	2DHIL(Partial)	87.45	87	87.88	91.32	90	87.65	90.67	88	92.11
	2DHIL	85.55	90	89.54	90.56	90	88.88	**92.56**	**91**	**93.5**
5x	2DHIL(Partial)	88.89	87	87.12	92.67	90	93.46	88.56	87	89.87
	2DHIL	90.76	90	91.23	**93.45**	**89**	**94.78**	88.96	87	88.89
10x	2DHIL(Partial)	90.54	90	94.76	90.56	89	91.22	89.88	87	88.78
	2DHIL	**91.78**	**93**	**95**	90.62	91	90.87	86.66	85	86.89

epochs at each incremental stage. We used Adam optimizer, a learning rate of 0.0006, and a dropout rate of 0.1 to prevent overfitting. Moreover, we split data with a ratio of 80,10 and 10 for forming training, validation, and test sets respectively. we created patches of 256×256 for our model. We also employed data augmentation by varying brightness, contrast, saturation, and hue. Furthermore, of all the resolutions available in the dataset we used 2x, 5x, and 10x resolutions for data incremental learning; the same resolutions were used for validation and testing.

5. Results and discussions

5.1. Initial results

In our experiments, we employ the Segformer-B0 model [44] as the backbone for the incremental semantic segmentation. The Segformer-B0 model is the lightest among different Segformer models with 3,319,292 tunable parameters.

We trained our Segformer model using training data of three different specifications. We used data of 10x, 5x, and 2x resolutions for training three Segformer models, initially pre-trained on ImageNet [75], thus we obtained three different trained models. We validated each of these three models using validation datasets corresponding to their respective resolutions, and subsequently tested them using data from all resolutions. We evaluated the performance of the proposed framework using three performance measures: accuracy, mean intersection over union (mIoU), and F1 score. The results of the experiment are listed in Table 1.

Table 1 indicates that the model trained on 10x data achieved the highest accuracy, mIoU, and F1 score, with values of 92.0, 89.0, and 90.89, respectively. The next best performance is demonstrated by the model trained on 5x resolution with an accuracy of 91.05, mIoU of 88.0, and F1 score of 89.56. The model trained on 2x resolution manifested the lowest performance with accuracy dropping to 90.98, mIoU to 86.0, and F1 score reaching 89.78. It is, however, important to notice that the F1 score for the least performing model is slightly higher than the model trained on 5x which may be because of the ways these different performance measures are calculated. It also highlights that the change in resolution may affect the performance measures vis-à-vis the overall proportion of classes in test samples. Table 1 further indicates a strong correlation between the accuracy of the model and the resolution of training samples. Additionally, it emphasizes that each model performs best when tested on samples of the same resolution it was trained on. As reflected in Table 1, for the model trained on 10x resolution, the best results are given by test samples of 10x. Similar is the case for the models trained on 5x and 2x resolutions. These results clearly indicate that the performance of a given model considerably deteriorates with the variation in the specification of test samples from the data on which the model is trained. In our case, the maximum decline in performance is observed when the model trained on 10x resolution is tested for test samples of 2x resolution manifesting a 6.03% decrease in accuracy where it reduces from 92.0% for 10x test cases to 86.45% for 2x test samples. This decrease highlights the lack of robustness in the segmentation model where a minor variation in the specifications of test samples from the training samples results in a 6.03% decline in accuracy. It is also important to note that the best results are given by the model trained on the data of the lowest resolution for the test samples of the same resolution. This suggests that enough information is available in samples of 10x resolution for the model to make predictions with 92% accuracy. Further experiments with data samples of different resolutions may indicate an optimal resolution of data for which our semantic segmentation model may achieve optimal accuracy. Consequently, preprocessing of training and test samples to optimal resolution may help in achieving better segmentation performance. We further present the visual results of our experiment corresponding to Table 1 in Fig. 4 giving test samples, masks, and corresponding predictions by each model. We find that the visual analysis of Fig. 4 strongly conforms to the analytical results given in Table 1.

5.2. Ablation studies

While implementing the 2DHIL framework as given in Fig. 1 using Segformer as a backbone, we approached the problem at hand in two different ways. Firstly we implemented the 2DHIL framework while freezing part of the layers of Segformer and allowed training in only five layers in each subsequent step; we name this model as 2DHIL(Partial). In the second attempt, we allowed training in all layers of the Segformer model while implementing the 2DHIL framework. We name this model as 2DHIL. The details of each model and corresponding results are discussed in succeeding paras.

5.2.1. 2DHIL(partial) - training partial layers

In this part of the ablation study, we implemented the 2DHIL framework with partial layer training, referred to as 2DHIL(Partial). Using this methodology, we trained the model in three incremental phases, using data of 2x, 5x, and then 10x resolution sequentially. Segformer was employed as the backbone of our framework. The results of this experiment are listed in Table 2. The 2DHIL(Partial) model was trained in three phases sequentially: first using 2x, then 5x, and finally with 10x data.

The best results were achieved by 2DHIL(Partial) at the culmination of training with 10x data, with an accuracy, mIoU, and F1 score of 94.76%, 90.00%, and 90.54%, respectively. We find that the next best result was achieved by 2DHIL(Partial) at the culmination of the second phase, where the model was trained on 5x data, attaining an accuracy of 93.46%. Additionally, 2DHIL(Partial) demonstrated the lowest accuracy of 92.11% after the first phase of training with 2x resolution data.

We observe that the accuracy of the model improved with incremental learning. The initial accuracy of 92.11% at the end of the first phase increased to 94.76% after the third phase, representing a 2.87% improvement. Furthermore, we observed a correlation between the resolution of the training data and the test data, as seen in the initial results. 2DHIL(Partial) gives the best results for test data that are similar to the respective training data.

Furthermore, we note that even with partial layer training, which reduces trainable parameters, following the proposed 2DHIL

Table 3
Comparison of segmentation accuracy among various incremental learning frameworks employing different loss functions and differently trained segmentation models. The accuracy has been calculated and listed separately for test samples of 2x, 5x, and 10x resolutions.

Test Data Resolution	Model	Fine-tuning	LwM [4]	L_{IL} [17]	L_{iss}
2x	Segformer	87.78	89.67	90.54	90.98
	2DHIL (Partial)	85.89	86.67	88.67	92.11
	2DHIL	79.16	80.82	84.45	**93.50**
5x	Segformer	89.67	90.32	91.78	91.05
	2DHIL (Partial)	87.88	89.90	89.76	93.46
	2DHIL	82.45	85.68	88.89	**94.78**
10x	Segformer	92.78	93.78	94.56	92.00
	2DHIL (Partial)	89.99	90.66	90.66	94.76
	2DHIL	88.76	89.98	90.14	**95.00**

methodology given in Fig. 1, the accuracy of 2DHIL(Partial) models improved across all three phases (resolutions) compared to batch training. Specifically, at the end of the incremental training in the third phase with 10x data, the accuracy improved by 3% from 92.00% to 94.76%. In the second phase, with training using 5x data, the increase was 2.64%, improving to 93.46% from 91.05%. Similarly, in the third phase, the accuracy increased by 1.24%.

5.2.2. 2DHIL - training all layers

Subsequently, we implemented the 2DHIL framework, allowing training in all layers. Here again, we initiated training in the initial phase with data of 2x resolution. Later, we incrementally implemented the 2DHIL framework by training with 5x resolution in the second phase and with 10x resolution in the third phase. The results of the phase-wise training of the 2DHIL framework are presented in Table 2, organized into different columns.

It is noteworthy that the 2DHIL replicated the same trend as in the case of 2DHIL(Partial). The highest accuracy is achieved at the culmination of training after the third phase with data of 10x resolution. The 2DHIL framework, after completion of training, achieved an accuracy, mIoU, and F1 score of 95.00%, 93.00, and 91.78, respectively. The next best accuracy of 94.78% was achieved after the second phase of training with data of 5x resolution, followed by the first phase, which achieved an accuracy of 93.5%. It is important to note that the improvement in accuracy for the 2DHIL model from the first phase to the final phase has been 1.6%, as the accuracy improved from 93.5% at the end of the first phase to 95% at the end of the third. These results clearly indicate the enhanced robustness of the model and its reduced sensitivity to data specifications.

Furthermore, it is evident that our proposed 2DHIIL framework achieves the best results, with a segmentation accuracy of 95.00%. Additionally, the variation in accuracy among test samples of different resolutions has decreased from 6.01% to 5.7%, as 10x test samples achieved an accuracy of 95.00% and 2x samples achieved 89.54% accuracy, demonstrating the improvement in the robustness of the proposed framework. It is also important to highlight that the proposed 2DHIL model demonstrated an improvement in overall accuracy as well. At the completion of incremental training in the third phase, it achieved an accuracy of 95.00%, compared to 92.00% in the case of batch training with the Segformer model (Table 1). In the second phase, using 5x data, the accuracy improved by 3.93% to 94.78%, compared to 91.05% in batch training. Similarly, in the first phase with 2x data, the accuracy increased by 2.69%.

5.3. Comparative results

We present the results of our proposed 2DHIL framework in Table 3 in comparison with other incremental learning methodologies. We tested fine-tuning, LwM [4] and L_{IL} [17] in comparison with our proposed framework for test samples of 2x, 5x, and 10x resolutions.

Table 4
Comparison of segmentation accuracy among various incremental learning frameworks employing different loss functions and differently trained segmentation models. The accuracy has been calculated for a test set containing samples from all classes and resolutions.

Test Data Resolution	Model	Fine-tuning	LwM [4]	L_{IL} [17]	L_{iss}
2x, 5x and 10x (Mixed and randomized)	Segformer	83.67	85.67	87.90	88.78
	2DHIL (Partial)	87.78	88.90	89.98	87.78
	2DHIL	86.56	87.89	88.89	**89.98**

As shown in Table 3, our proposed 2DHIL framework while employing the loss function L_{iss} given in Eq. (1), performs significantly better than well-known methodologies for test samples of all resolutions. For 10x resolution test samples, 2DHIL achieved 95.00% accuracy which is over 6% more than its nearest competitor L_{IL} [17] that achieved an accuracy of 90.14%. Similarly, 2DHIL performed 7.45% better in the case of 5x and about 11.6% better in the case of 2x test samples. It is also noticeable that the 2DHIL(Partial) framework although showed less performance as compared to the 2DHIL framework, its performance has been superior to the rest of the competitor benchmarks in test samples of all resolutions. It is, however, questionable that for the standard Segformer model, two loss functions L_{IL} and L_{iss} being very similar manifested different results, and this difference is particularly pronounced in the case of 10x where it reaches to 3.3%. To our analysis, this difference in performance may be due to the difference in the proportion of data belonging to new and old classes. We further compared the performance of our proposed framework with other methodologies with a test set having test samples of all resolutions. The results of this experiment are listed in Table 4 which shows that the proposed 2DHIL framework exhibits the best performance by achieving an accuracy of 89.98% for combined samples of all resolutions and all carcinomas. The improved performance demonstrated by the proposed framework for test samples of varying specifications indicates improvement in the robustness of the proposed framework. It is also noticeable that the 89.98% accuracy achieved by 2DHIL for test data of varying resolutions is better than all those cases given in Table 1 where test data is different from training data. We also present the visual results of our proposed model corresponding to Table 4 in Fig. 5 which shows an improvement in predictions particularly in cases of 5x and 2x test samples.

To further emphasize the efficacy of the proposed 2DHIL framework, we trained six different segmentation models employing the L_{iss} loss function and tested them separately with test samples of 2x, 5x, and 10x resolutions. The results of the experiment are listed in Table 5. These results highlight that the best segmentation accuracy is achieved by the 2DHIL segmentation framework (highlighted in bold) compared to all other segmentation models. Additionally, despite part of the 2DHIL (Partial) model being frozen, it manifested the second-best performance (underlined). In conclusion, our results demonstrate the effectiveness of the 2DHIL framework in achieving high segmentation accuracy across different resolutions, highlighting its potential for robust and adaptable segmentation in medical imaging.

We, moreover, present a classwise accuracy comparison in Table 6 and Fig. 6 accounting for all the twelve classes including background. As reflected in Table 6, the 2DHIL framework outperformed benchmark work [69] using the same data. Out of all twelve classes, the 2DHIL framework performed better in nine classes. The accuracy improvement has been between 3% to 54% whereas for six classes the percentage of improvement has been in double figures. The best improvement of 54% is noticeable for INF where the model achieved an accuracy of 88% in comparison with 57% of [69]. The considerable improvement in classification accuracy for INF may be attributed to the provision of training

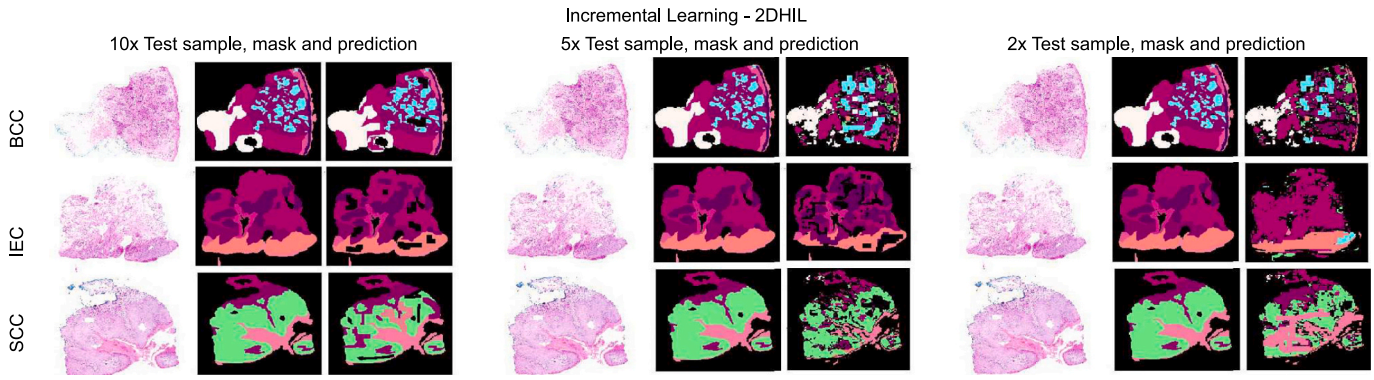

Fig. 5. Multi-resolution samples of non-melanoma skin cancers (BCC, IEC, and SCC), corresponding segmentation masks and predictions by 2DHIL framework.

Table 5
Comparison of segmentation accuracy achieved by different models with test samples of varying resolutions.

	Average Accuracy		
Model	Test Data 10x	Test Data 5x	Test Data 2x
SegNet [38]	90.12	88.29	86.31
FCN [33]	86.31	84.48	82.50
PSPNet [41]	89.92	88.09	86.11
U-Net [34]	90.17	88.34	86.35
LW-Net	90.97	89.53	88.32
Segformer [44]	92.00	91.05	90.98
2DHIL(Partial)	94.76	93.46	92.11
2DHIL	**95.00**	**94.78**	**93.5**

data of multiple resolutions to the model which is analogous to the way pathologists zoom in or zoom out their scopes during diagnosis. It is also noticeable that the proposed framework under-performed for GLD and EPI where the performance degraded by 2% and 7% respectively. The lowest accuracy achieved by the proposed framework is for EPI which is about 77%. It is logical to compare that [69] classified four classes i.e. IEC, INF, RET, and FOL with accuracy lower than this. It is also encouraging to notice that the proposed framework particularly identified all carcinomas i.e. BCC, SCC, and IEC with 92%, 96%, and 92% accuracy manifesting an improvement of 7%, 13%, and 31% respectively in comparison with [69].

It is important to highlight that the proposed framework achieved the leading segmentation accuracy of 98% for BKG, followed by 97% for RET and 96% for SCC and HYP. The superior segmentation performance for these classes can be attributed to their higher representation in the data as BKG constitutes the largest portion of the data accounting for

Table 6
Classwise segmentation accuracy of the comparative models. The highest and the next best accuracy achieved by a model for each class is highlighted as bold and underlined respectively. The percentage difference in accuracy is the difference from S.M. Thomas et al. [69] in comparison with the 2DHIL framework where all positive values highlighted in bold indicate enhanced accuracy achieved by the proposed 2DHIL framework.

Model	BKG	BCC	SCC	IEC	EPI	GLD	INF	RET	FOL	PAP	HYP	KER
S.M. Thomas et al. [69]	0.95	0.86	0.85	0.70	**0.83**	**0.87**	0.57	0.70	0.61	0.80	**0.96**	0.84
U-Net [34]	**0.99**	0.91	0.70	0.82	0.70	0.81	0.64	0.91	0.65	0.68	0.85	0.79
2DHIL	0.98	**0.92**	**0.96**	**0.92**	0.77	0.85	**0.88**	**0.97**	**0.83**	**0.89**	**0.96**	**0.90**
Percentage difference	**3%**	**7%**	**13%**	**31%**	−7%	−2%	**54%**	**39%**	**36%**	**11%**	0%	**7%**

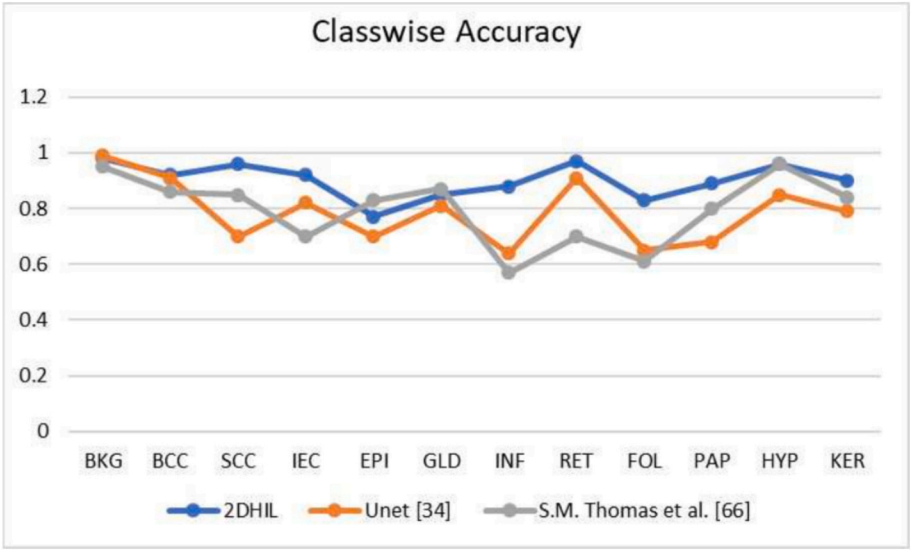

Fig. 6. Classwise accuracy comparison among 2DHIL, U-Net [34], and S.M. Thomas et al. [69].

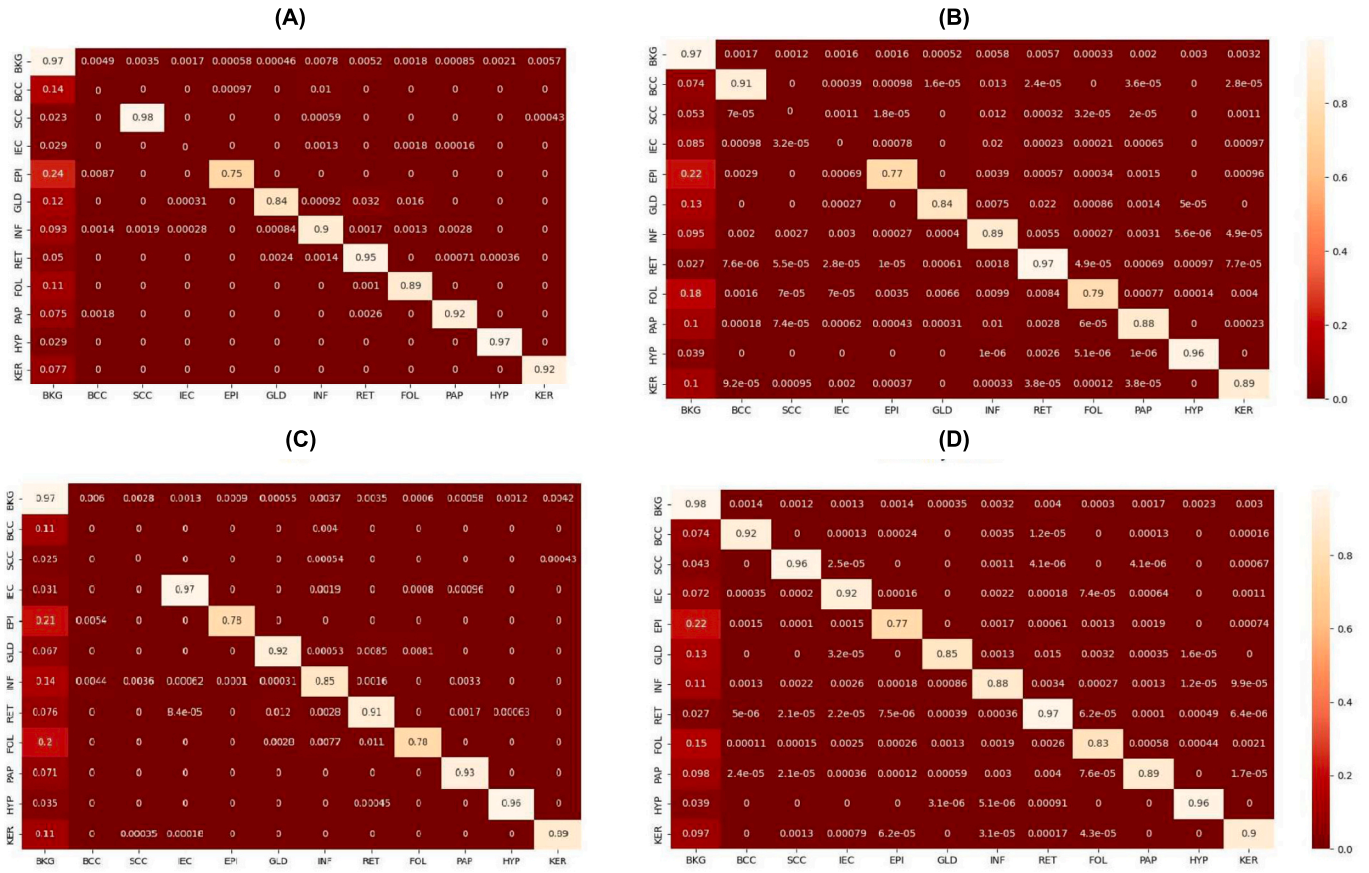

Fig. 7. Confusion matrices of predictions by the 2DHIL framework. (A) Confusion matrix for SCC test cases. (B) Confusion matrix for BCC test cases. (C) Confusion matrix for IEC test cases. (D) Confusion matrix for all test cases.

58.6% of all pixels, followed by RET (19.8%), HYP (9.3%), and INF (2.8%). Furthermore, among carcinomas, the best performance is demonstrated for SCC with an accuracy of 96%, while BCC and IEC both trail behind at 92%. It is worth noting that SCC has the highest pixel representation among carcinomas with 2.4% of pixels, whereas the representation of BCC and IEC is 1.1% and 1.0% respectively. Consequently, the performance of BCC and IEC is lower as compared to SCC and other classes having higher representation in data.

The confusion matrices indicating classwise accuracy achieved by the proposed model corresponding to test samples of different carcinomas are presented in Fig. 7.

6. Conclusion

Acquiring the ability to evolve and grow with time is the biggest challenge for AI models. Incremental learning, however, can provide these models with the capability of evolution and enhanced robustness through learning in sequential steps. The two-dimensional hybrid incremental learning framework for multi-class semantic segmentation, introduced in this study, demonstrates its successful implementation in a complex problem of identifying carcinomas and skin tissues from histopathology slides. The results of our extended experiments indicate that after incrementally training through our proposed framework an AI model, Segformer in our case, becomes adaptable and can be evolved to segment more and more classes. It is important to note that the model also evolves to process data of varying specifications and thus by minimizing its sensitivity to the specification of data it becomes robust and capable of processing test cases of different specifications with high prediction accuracy. The successful segmentation of 12 classes with high accuracy by the proposed framework using a very limited number of training samples at each incremental learning stage demonstrates the efficacy of our work. It is also noticeable that catastrophic forgetting has been addressed mainly through a combination of knowledge distillation and replay of old classes without adding any memory component to the model. The proposed framework does not require any additional components to learn new classes however since the model has fixed tunable parameters therefore a stage of saturation may come after multiple cycles of incremental training in cases where new classes are continuously added without dropping any older classes. In the case of an application where some classes become obsolete over time on the emergence of new classes, the model may keep adapting indefinitely and the stage of saturation may come much later.

It is important to note that while implementing the 2DHIL framework, we incrementally presented data in the order of 2x, 5x, and 10x in three different phases. However, we have not evaluated the effect of presenting data in different orders. Future studies should explore the impact of different resolution orders in incremental learning frameworks. By investigating sequences like 5x, 2x, 10x, and others, researchers can optimize strategies for handling evolving datasets, enhancing model adaptability in real-world scenarios. This exploration will provide valuable insights for refining incremental learning methodologies. Furthermore, expanding the evaluation of the proposed framework to include various AI models beyond Segformer could validate its versatility and applicability across different architectures. The application of the proposed framework in real-life scenarios in the future could help in further improvements.

References

[1] R.M. French, Catastrophic forgetting in connectionist networks, Trends Cogn. Sci. 3 (4) (1999) 128–135.

[2] M. McCloskey, N.J. Cohen, Catastrophic interference in connectionist networks: The sequential learning problem, in: G.H. Bower (Ed.), Ser. Psychology of Learning and Motivation vol. 24, Academic Press, 1989, pp. 109–165 [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0079742108605368.

[3] Z. Li, D. Hoiem, Learning without forgetting, IEEE Trans. Pattern Anal. Mach. Intell. 40 (12) (2018) 2935–2947.

[4] P. Dhar, R.V. Singh, K. Peng, Z. Wu, R. Chellappa, Learning without memorizing, in: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019.

[5] S. Rebuffi, A. Kolesnikov, G. Sperl, C.H. Lampert, Icarl: Incremental classifier and representation learning, in: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE Computer Society, Los Alamitos, CA, USA, Jul 2017, pp. 5533–5542 [Online]. Available: https://doi.org/10.1109/CVPR.2017.587.

[6] H. Jung, J. Ju, M. Jung, J. Kim, Less-forgetting learning in deep neural networks, arXiv (2016) preprint arXiv:1607.00122.

[7] R. Aljundi, F. Babiloni, M. Elhoseiny, M. Rohrbach, T. Tuytelaars, Memory aware synapses: Learning what (not) to forget, in: Proceedings of the European Conference on Computer Vision (ECCV), September 2018.

[8] A. Chaudhry, P.K. Dokania, T. Ajanthan, P.H.S. Torr, Riemannian walk for incremental learning: Understanding forgetting and intransigence, in: Proceedings of the European Conference on Computer Vision (ECCV), September 2018.

[9] J. Kirkpatrick, R. Pascanu, N. Rabinowitz, J. Veness, G. Desjardins, A.A. Rusu, K. Milan, J. Quan, T. Ramalho, A. Grabska-Barwinska, et al., Overcoming catastrophic forgetting in neural networks, Proc. Natl. Acad. Sci. 114 (13) (2017) 3521–3526.

[10] D. Lopez-Paz, M.A. Ranzato, Gradient episodic memory for continual learning, in: I. Guyon, U.V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, R. Garnett (Eds.), Advances in Neural Information Processing Systems vol. 30, Curran Associates, Inc., 2017 [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2017/file/f87522788a2be2d171666752f97ddebb-Paper.pdf.

[11] U. Michieli, P. Zanuttigh, Incremental learning techniques for semantic segmentation, in: Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops, 2019.

[12] U. Michieli, Zanuttigh, Continual semantic segmentation via repulsion-attraction of sparse and disentangled latent representations, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 1114–1124.

[13] A. Douillard, Y. Chen, A. Dapogny, M. Cord, Plop: Learning without forgetting for continual semantic segmentation, in: 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), IEEE Computer Society, Los Alamitos, CA, USA, Jun 2021, pp. 4039–4049 [Online]. Available: https://doi.org/10.1109/CVPR46437.2021.00403.

[14] F. Cermelli, M. Mancini, S.R. Bulo, E. Ricci, B. Caputo, Modeling the background for incremental learning in semantic segmentation, in: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), IEEE Computer Society, Los Alamitos, CA, USA, Jun 2020, pp. 9230–9239 [Online]. Available: https://doi.org/10.1109/CVPR42600.2020.00925.

[15] S. Cha, B. Kim, Y. Yoo, T. Moon, Ssul: Semantic segmentation with unknown label for exemplar-based class-incremental learning, in: M. Ranzato, A. Beygelzimer, Y. Dauphin, P. Liang, J.W. Vaughan (Eds.), Advances in Neural Information Processing Systems vol. 34, Curran Associates, Inc., 2021, pp. 10919–10930 [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2021/file/5a9542c773018268fc6271f7afeea969-Paper.pdf.

[16] G. Hinton, O. Vinyals, J. Dean, Distilling the knowledge in a neural network, arXiv (2015) preprint arXiv:1503.02531.

[17] A.M. Khan, T. Hassan, M.U. Akram, N.S. Alghamdi, N. Werghi, Continual learning objective for analyzing complex knowledge representations, Sensors 22 (4) (2022) [Online]. Available: https://www.mdpi.com/1424-8220/22/4/1667.

[18] M. Sirshar, T. Hassan, M.U. Akram, S.A. Khan, An incremental learning approach to automatically recognize pulmonary diseases from the multi-vendor chest radiographs, Comput. Biol. Med. 134 (2021) 104435 [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0010482521002298.

[19] U. Michieli, P. Zanuttigh, Knowledge distillation for incremental learning in semantic segmentation, Comput. Vis. Image Underst. 205 (2021) 103167.

[20] Y. Tian, D. Krishnan, P. Isola, Contrastive representation distillation, arXiv (2019) preprint arXiv:1910.10699.

[21] F.M. Castro, M.J. Marín-Jiménez, N. Guil, C. Schmid, K. Alahari, End-to-end incremental learning, in: Proceedings of the European Conference on Computer Vision (ECCV), 2018, pp. 233–248.

[22] E. Belouadah, A. Popescu, Il2m: Class incremental learning with dual memory, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019, pp. 583–592.

[23] S. Mittal, S. Galesso, T. Brox, Essentials for class incremental learning, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 3513–3522.

[24] D. Lopez-Paz, M. Ranzato, Gradient episodic memory for continual learning, Adv. Neural Inf. Proces. Syst. 30 (2017).

[25] C. Zhuang, S. Huang, G. Cheng, J. Ning, Multi-criteria selection of rehearsal samples for continual learning, Pattern Recogn. 132 (2022) 108907.

[26] J. Smith, Y.-C. Hsu, J. Balloch, Y. Shen, H. Jin, Z. Kira, Always be dreaming: A new approach for data-free class-incremental learning, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 9374–9384.

[27] A. Maracani, U. Michieli, M. Toldo, P. Zanuttigh, Recall: Replay-based continual learning in semantic segmentation, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 7026–7035.

[28] Y. Yang, B. Chen, H. Liu, Bayesian compression for dynamically expandable networks, Pattern Recogn. 122 (2022) 108260 [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0031320321004404.

[29] S. Yan, J. Xie, X. He, Der: Dynamically expandable representation for class incremental learning, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 3014–3023.

[30] R. Aljundi, P. Chakravarty, T. Tuytelaars, Expert gate: Lifelong learning with a network of experts, in: Proceedings of the IEEE Conference on Computer Vision and pattern recognition, 2017, pp. 3366–3375.

[31] R. Caruana, Multitask learning, Mach. Learn. 28 (1997) 41–75.

[32] G.M. van de Ven, T. Tuytelaars, A.S. Tolias, Three types of incremental learning, Nat. Machine Intellig. 4 (12) (2022) 1185–1197.

[33] J. Long, E. Shelhamer, T. Darrell, Fully convolutional networks for semantic segmentation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 3431–3440.

[34] O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, in: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18, Springer, 2015, pp. 234–241.

[35] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, A.L. Yuille, Deeplab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs, IEEE Trans. Pattern Anal. Mach. Intell. 40 (4) (2018) 834–848.

[36] F.S.H.A. Liang-Chieh Chen, George Papandreou, Rethinking atrous convolution for semantic image segmentation, CoRR abs/1706.05587 (2017) [Online]. Available: http://arxiv.org/abs/1706.05587.

[37] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, H. Adam, Encoder-decoder with atrous separable convolution for semantic image segmentation, in: Proceedings of the European Conference on Computer Vision (ECCV), 2018, pp. 801–818.

[38] V. Badrinarayanan, A. Kendall, R. Cipolla, Segnet: a deep convolutional encoder-decoder architecture for image segmentation, IEEE Trans. Pattern Anal. Mach. Intell. 39 (12) (2017) 2481–2495.

[39] A. Paszke, A. Chaurasia, S. Kim, E. Culurciello, Enet: A deep neural network architecture for real-time semantic segmentation, arXiv (2016) preprint arXiv:1606.02147.

[40] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770–778.

[41] H. Zhao, J. Shi, X. Qi, X. Wang, J. Jia, Pyramid scene parsing network, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 2017.

[42] M. Tan, Q. Le, Efficientnet: Rethinking model scaling for convolutional neural networks, in: International Conference on Machine Learning, PMLR, 2019, pp. 6105–6114.

[43] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, et al., An image is worth 16x16 words: Transformers for image recognition at scale, arXiv (2020) preprint arXiv:2010.11929.

[44] E. Xie, W. Wang, Z. Yu, A. Anandkumar, J.M. Alvarez, P. Luo, Segformer: Simple and efficient design for semantic segmentation with transformers, in: Advances in Neural Information Processing Systems vol. 34, 2021, pp. 12077–12090.

[45] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, B. Guo, Swin transformer: Hierarchical vision transformer using shifted windows, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 10012–10022.

[46] S. Zheng, J. Lu, H. Zhao, X. Zhu, Z. Luo, Y. Wang, Y. Fu, J. Feng, T. Xiang, P. H. Torr, L. Zhang, Rethinking semantic segmentation from a sequence-to-sequence perspective with transformers, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 2021, pp. 6881–6890.

[47] B. Cheng, A. Schwing, A. Kirillov, Per-pixel classification is not all you need for semantic segmentation, in: Advances in Neural Information Processing Systems vol. 34, 2021, pp. 17864–17875.

[48] J. Chen, Y. Lu, Q. Yu, X. Luo, E. Adeli, Y. Wang, L. Lu, A.L. Yuille, Y. Zhou, Transunet: Transformers make strong encoders for medical image segmentation, arXiv (2021) preprint arXiv:2102.04306.

[49] W. Wang, E. Xie, X. Li, D.-P. Fan, K. Song, D. Liang, T. Lu, P. Luo, L. Shao, Pyramid vision transformer: A versatile backbone for dense prediction without convolutions, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 568–578.

[50] D. Zhao, B. Yuan, Z. Shi, Inherit with distillation and evolve with contrast: exploring class incremental semantic segmentation without exemplar memory, in: IEEE Transactions on Pattern Analysis & Machine Intelligence 45, Oct 2023, pp. 11932–11947, no. 10.

[51] U. Michieli, P. Zanuttigh, Incremental learning techniques for semantic segmentation, in: Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops, 2019.

[52] Y. Oh, D. Baek, B. Ham, Alife: Adaptive logit regularizer and feature replay for incremental semantic segmentation, in: Advances in Neural Information Processing Systems vol. 35, 2022, pp. 14516–14528.

[53] T. Kalb, J. Beyerer, Causes of catastrophic forgetting in class-incremental semantic segmentation, in: Proceedings of the Asian Conference on Computer Vision, 2022, pp. 56–73.

[54] C.-B. Zhang, J.-W. Xiao, X. Liu, Y.-C. Chen, M.-M. Cheng, Representation compensation networks for continual semantic segmentation, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 7053–7064.

[55] L. Yu, X. Liu, J. Van de Weijer, Self-training for class-incremental semantic segmentation, in: IEEE Transactions on Neural Networks and Learning Systems, 2022.
[56] D. Goswami, R. Schuster, J. van de Weijer, D. Stricker, Attribution-aware weight transfer: A warm-start initialization for class-incremental semantic segmentation, in: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2023, pp. 3195–3204.
[57] F. Cermelli, M. Mancini, S.R. Bulo, E. Ricci, B. Caputo, Modeling the background for incremental and weakly-supervised semantic segmentation, IEEE Trans. Pattern Anal. Mach. Intell. 44 (12) (2021) 10099–10113.
[58] Y. Yang, S. Yuan, L. Xie, Overcoming catastrophic forgetting for semantic segmentation via incremental learning, in: 2022 17th International Conference on Control, Automation, Robotics and Vision (ICARCV), IEEE, 2022, pp. 299–304.
[59] Y. Qiu, Y. Shen, Z. Sun, Y. Zheng, X. Chang, W. Zheng, R. Wang, Sats: self-attention transfer for continual semantic segmentation, Pattern Recogn. 138 (2023) 109383.
[60] C. Shang, H. Li, F. Meng, Q. Wu, H. Qiu, L. Wang, Incrementer: Transformer for class-incremental semantic segmentation with knowledge distillation focusing on old class, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023, pp. 7214–7224.
[61] R. Strudel, R. Garcia, I. Laptev, C. Schmid, Segmenter: Transformer for semantic segmentation, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 7262–7272.
[62] K. Li, L. Yu, P.-A. Heng, Domain-incremental cardiac image segmentation with style-oriented replay and domain-sensitive feature whitening, IEEE Trans. Med. Imaging 42 (3) (2023) 570–581.
[63] W. He, X. Wang, L. Wang, Y. Huang, Z. Yang, X. Yao, X. Zhao, L. Ju, L. Wu, L. Wu, H. Lu, Z. Ge, Incremental learning for exudate and hemorrhage segmentation on fundus images, Inform. Fusion 73 (2021) 157–164 [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1566253521000427.
[64] M.S. Arifin, M.G. Kibria, A. Firoze, M.A. Amini, H. Yan, Dermatological disease diagnosis using color-skin images, in: 2012 International Conference on Machine Learning and Cybernetics vol. 5, IEEE, 2012, pp. 1675–1680.
[65] P.M. Pereira, R. Fonseca-Pinto, R.P. Paiva, P.A. Assuncao, L.M. Tavora, L. A. Thomaz, S.M. Faria, Dermoscopic skin lesion image segmentation based on local binary pattern clustering: comparative study, Biomed. Sign. Proc. Control 59 (2020) 101924.
[66] F. Afza, M. Sharif, M.A. Khan, U. Tariq, H.-S. Yong, J. Cha, Multiclass skin lesion classification using hybrid deep features selection and extreme learning machine, Sensors 22 (3) (2022) 799.
[67] M.H. Hesamian, W. Jia, X. He, P. Kennedy, Deep learning techniques for medical image segmentation: achievements and challenges, J. Digit. Imaging 32 (2019) 582–596.
[68] A. Bibi, M.A. Khan, M.Y. Javed, U. Tariq, B.-G. Kang, Y. Nam, R.R. Mostafa, R. H. Sakr, Skin lesion segmentation and classification using conventional and deep learning based framework, Comput. Mater. Contin 71 (2) (2022) 2477–2495.
[69] S.M. Thomas, J.G. Lefevre, G. Baxter, N.A. Hamilton, Interpretable deep learning systems for multi-class segmentation and classification of non-melanoma skin cancer, Med. Image Anal. 68 (2021) 101915 [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1361841520302796.
[70] S. Öztürk, U. Özkaya, Skin lesion segmentation with improved convolutional neural network, J. Digit. Imaging 33 (2020) 958–970.
[71] H. Wu, J. Pan, Z. Li, Z. Wen, J. Qin, Automated skin lesion segmentation via an adaptive dual attention module, IEEE Trans. Med. Imaging 40 (1) (2020) 357–370.
[72] P. Kleczek, J. Jaworek-Korjakowska, M. Gorgon, A novel method for tissue segmentation in high-resolution h&e-stained histopathological whole-slide images, Comput. Med. Imaging Graph. 79 (2020) 101686.
[73] H. Wu, S. Chen, G. Chen, W. Wang, B. Lei, Z. Wen, Fat-net: feature adaptive transformers for automated skin lesion segmentation, Med. Image Anal. 76 (2022) 102327.
[74] S.M. Thomas, J.G. Lefevre, G. Baxter, N.A. Hamilton, Non-melanoma skin cancer segmentation for histopathology dataset, Data Brief 39 (2021) 107587 [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2352340921008623.
[75] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, L. Fei-Fei, Imagenet: A large-scale hierarchical image database, in: 2009 IEEE Conference on Computer Vision and Pattern Recognition, 2009, pp. 248–255.