Yuty's Project: Detection of Multiple Facial Skin Concerns using Advanced ResNet50 U-Net Model

MSc Business Analytics 2022/2023

BUSM130 Group 9

Saleha Zahid: 220691486
Senkun Xiang: 220967130
Snehal Srivastava: 220316567
Zhuoxin Ye: 220723020
Xi Ye: 200630355

Supervisors

Chris Sutton, Lei Fang, Natalia Efremova

AUGUST 21, 2023

Student's Declaration

We,

Saleha Zahid: 220691486

Senkun Xiang: 220967130

Snehal Srivastava: 220316567

Zhuoxin Ye: 220723020

Xi Ye: 200630355

hereby declare that the work in this Group Project Report is our original work. We have not copied from any other students' work, work of ours submitted elsewhere, or from any other sources except where due reference or acknowledgement is made explicitly in the text, nor has any part been written for us by another person. The individual sections of the report have been written by the author in the section title with no external assistance.

Referenced text has been flagged by:

- Using italic fonts, **and**
- using quotation marks " ", **and**
- explicitly mentioning the source in the text.

Table of Contents

Table of Figures

Tables

1. Executive Summary

1.1. Background

The beauty industry is a thriving realm projected to achieve over $101 billion in revenue by 2027, encompassing iconic brands such as L'Oréal and Sephora (Howarth, 2023). Amidst a staggering array of product offerings, customers encounter difficulty in selecting suitable skincare solutions. Personalization emerges as a pivotal factor influencing customer loyalty and repeat purchases. Yuty, a small yet visionary beauty enterprise, embarks on a mission to alleviate this challenge through the ingenious application of AI and machine learning, culminating in personalized beauty experiences.

In the sphere of personalized skincare, our literature review illuminates the pivotal role of image segmentation. We navigate through fundamental segmentation types: binary and semantic, hinging on established techniques such as thresholding and edge detection. Notably, deep learning architectures like FCNs, U-Net, and Mask R-CNN stand out for their semantic and instance segmentation prowess. Among the mentioned architectures, U-net stands as a pivotal framework in CNN-driven semantic segmentation, frequently leveraged in medical imaging to segment lesions, delineate anatomical structures, and classify anomalies (Du, et al 2020), also achieved decent result in a variety of biomedical segmentation applications (Ronneberger, Fischer & Brox, 2015).

While U-Net has been widely acknowledged for its efficacy in the medical sector, its application in the beauty industry still needs to be explored, presenting an intriguing gap. Our study integrating U-net with automatic facial skin concern detection by selfie images serves as a valuable reference for the beauty sector, drawing lessons from advancements in medical imaging. Moreover, by addressing the dilemma of product selection faced by consumers, we aim to empower cosmetic brands with innovative solutions, ultimately boosting customer satisfaction.

1.2. Business Goal

By developing a system that can accurately identify facial skin imperfections from users selfies, Yuty intends to harness the potential of image segmentation technology and aims to deploy a robust model capable of detecting nine skin concerns from customer images. Integrating such a model would enable Yuty to offer personalized skincare advice and provide virtual skincare product try-ons, facilitate remote dermatology consultations, and maintain a skincare progress tracker for customers. This strategic move is anticipated to significantly elevate customer experience, satisfaction, and loyalty, positioning Yuty at a vantage point in the competitive beauty industry landscape.

1.3. Research Objectives

Our study seeks to design and establish a U-Net deep learning model proficient in analyzing facial selfies and the associated metadata indicating skin issues. The model's objectives are threefold:

- Attain effective accuracy in recognizing nine skin concerns.
- Ensure model's efficiency to processes user’s selfies in real-time.
- Cater to a wide range of demographics and skin types, factoring in variations such as lighting, skin tones, and facial expressions.

Within the scope of this study, we delve into several pivotal research questions to optimize our machine learning model for skin issue detection: (1) whether U-Net and its variant is suitable for skin concern detection; (2) identification of other possible architectures choices; (3) the impact of data size on model performance; (4) solutions to address imbalanced datasets for under-represented skin concerns; (5) solutions to address imbalance between skin concern area versus the non-skin concern areas; (6) impact of ensemble learning on model enhancement.

1.4. Methodology

In our study, the U-Net convolutional neural network was employed to detect facial skin anomalies from selfie images. Our approach encompassed an in-depth comparison of eight experimental groups, leading us to discern the model best-suited for subsequent phases. U-Net's efficiency with limited datasets, facilitating swift training, was especially advantageous within Google Colab's computational constraints. This efficiency was magnified by pairing U-Net with encoders, notably VGG16, ResNet50, and EfficientNetB4, capitalizing on the merits of pre-trained models.

Subsequently, a meticulous comparison of eight experimental groups between six U-Net architecture and encoder groups was undertaken, indicating that both ResNet50 and EfficientNetB4 exhibited comparable strengths. While the Resnet50 model displayed robust learning capabilities on the training dataset and EfficientNetB4 stood out for higher validation accuracy, respectively. Hence, we further evaluated the two models with an expanded dataset from 327 to 1000. As a result, ResNet50's performance markedly surpassed that of EfficientNet across numerous skin attributes. ResNet50's mean IoU reached 0.38, outperforming EfficientNet's 0.32, reflecting its broader adaptability to diverse skin attributes. Hence, we endorsed ResNet50 as the primary encoder for ensuing model enhancements.

1.5. Key Findings

In the advanced stage of our research, we intensified efforts to ameliorate the prevailing overfitting predicament and elevate the overall model performance by employing eight distinct enhancement techniques to fine-tune the ResNet50 based U-Net model. Specifically, our model excels in detecting prominent skin concerns, especially Wrinkles Fine Lines, Redness, and Dark Circles, and has demonstrated a **14%** improvement in IoU compared to its initial version.

The eight optimizing strategies including dataset expansion, hyperparameter tuning, L2 regularization, Dropout implementation, decoder streamlining, a synergistic application of metrics including Accuracy, Dice Coefficient, and Recall; threshold calibrations; ensemble learning and facial area detections. During the process, we have also found answers for our research questions.

From our results, a significant impact was noted when enriching our dataset. An increase in prediction accuracy for skin concerns such as Redness, Dark Circles, and Wrinkles Fine Lines was observable with an expanded data volume. Notably, our model exhibited proficiency in identifying these skin concerns. Yet, a challenge persisted in detecting HDM (Hyperpigmentation Dark Marks) and Acne, with a recorded 0% accuracy rate. A deep dive revealed an underrepresentation of these classes in the training dataset. For instance, among 22,132 images of Yuty's database, Acne and HDM only constituted 5.8% and 7.1% respectively.

In addressing overfitting, hyperparameter optimization played a pivotal role. Techniques like L2 Regularization (Ridge Regression) and Dropout were deployed. Our findings indicated an optimal L2 value of 0.0001 for the validation set, enhancing model robustness. On the other hand, the Dropout technique curbed overfitting without impacting the validation dataset's results significantly.

However, the glaring data imbalance for classes, especially Acne and HDM, warranted innovative interventions. Our response was the integration of ensemble learning, harnessing different data subsets, and formulating a dedicated model for Acne prediction. This dual approach yielded a notable surge in prediction accuracy from 0% to 27% for predicting the under-represented classes. To fine-tune the model further, facial region detection was synergized, ensuring predictions homed in on facial intricacies, excluding extraneous backgrounds. This enhancement contributed to an IoU increment of 0.07%.

Conclusively, our refined model exhibited a 14% enhancement in IoU compared to its initial iteration. While the model effectively detects major skin concerns, notably Wrinkles Fine Lines, Redness, and Dark Circles. From a commercial standpoint, accurately identifying these prevalent skin concerns aligns with consumers' primary grievances. Being adept at accurately pinpointing these issues translates to the potential of delivering precise skincare recommendations, thereby augmenting customer satisfaction. Future endeavors should target areas such as Acne and HDM to ensure an all-inclusive skin analysis tool. This continued commitment will further solidify its practical application, ensuring accurate and personalized skincare recommendations.

1.6. Limitations & Future Development

At the heart of Yuty's commitment to superior skincare solutions lies our inventive ResNet 50-based U-net model, meticulously designed for the segmentation of nine specific skin concerns. While showcasing proficiency in identifying certain skin issues, this state-of-the-art architecture acknowledges the challenges inherent to pioneering technologies.

The journey in model development faced constraints, notably with the Google Colab platform restricting the volume of training data. Furthermore, the underrepresentation of particular skin concerns, like Acne and Hyperpigmentation Dark Marks, impacted the model's overarching predictive accuracy. Properly distributing a dataset to ensure equitable representation of all skin concerns, especially when one image might host multiple concerns, was a substantial logistical hurdle. An intrinsic limitation was the model's adaptability, where the restricted sample size potentially affected its sensitivity to diverse lighting, skin tones, and facial expressions. On a more advanced front, the evolving nature of technology required to discern subtle skin issues meant there is much ground yet to be covered in research. And lastly, the precision of the model at times was compromised by annotations that were noisy-labelled.

However, where challenges lie, opportunities for growth and refinement abound. We envision the utilization of high-performance clusters like Summit or Sierra to manage larger datasets. This will significantly expand our model's learning capacity. Actively sourcing richer datasets and potentially harnessing pre-trained models can provide the model with a more holistic understanding of the underrepresented skin concerns. We are also keenly exploring alternative architectures such as the Multi-scale Dense U-Net and the Attention U-Net, aiming to strike a balance between model complexity and its adaptability.

In summary, our efforts in refining Yuty's facial image segmentation model have been marked by challenges, but these hurdles highlight our deep commitment to enhancing the model's capabilities. Each obstacle has only deepened our knowledge and determination, reminding us that there's always room for improvement. The journey has shown us the immense potential in this field. At Yuty, we're not just content with what we've achieved; we're eager and prepared to invest even more efforts to push the boundaries of what our model can do for personalized skincare.

1.7. Business Implications

In today's dynamic beauty landscape, consumers grapple with the vast array of product options, reflecting the intricate matrix of the industry. Such diversification, while being a testament to innovation, often engenders the 'paradox of choice,' leaving consumers bewildered about their best fit. This evolving quandary accentuates the pressing need for personalization in beauty and skincare. Current trends underscore this sentiment, with data pointing towards a robust demand for bespoke solutions in the sector.

As a beacon of innovation, Yuty aligns its objectives to this demand, endeavouring to simplify the shopping journey using artificial intelligence to provide personalized services and elevate the consumer experience. With the established multi-facial concern detection model, Yuty could conduct a nuanced diagnosis of nine skin concerns, aiming to empower consumers with precise diagnoses and curated product suggestions. Such endeavours are not just novelty-driven but also align with tangible business outcomes. Research by Forrester underscores this, indicating that personalization via AI can precipitate a sales uplift of up to 10%.

Analyzing Yuty from a SWOT framework provides a clear roadmap for our project's potential. Yuty's strengths lie in its avant-garde positioning and commitment to AI-powered solutions, a niche that our project is primed to enhance. Our initiative will not only augment Yuty's tailored offerings but also elevate user contentment, stimulate repeat purchases, and streamline customer support expenditures. Nevertheless, Yuty's embryonic brand presence and limited consumer reach remain its Achilles' heel. Therefore, our project could serve as a linchpin, positioning Yuty at the forefront of AI-driven beauty solutions, attracting a more extensive consumer base, and reinforcing brand credibility.

Moreover, as the personalized beauty industry thrives, Yuty stands at the cusp of vast opportunities. However, the path is not without its shadows. Competition from established players, emerging contenders with aligned ambitions, and the ever-present concerns around data security and user privacy present formidable challenges. Our mandate extends beyond amplifying Yuty's core strengths; it encompasses strategizing through these complexities, ensuring that Yuty consistently outpaces its competitors and preemptively addresses potential privacy concerns.

In essence, Yuty's voyage towards market dominance, fortified by our AI solution, promises a sustainable combination of personalized excellence and business growth in the industry.

1.8. Team Work Reflection

Our journey with the Yuty project was chosen out of genuine interest. Snehal and Saleha, with their analytical minds, delved into research and business strategy, while Xi Ye, Senkun Xiang and Zhuoxin Ye, with their technical prowess, anchored the coding part. We leveraged Teams for communication, centralizing our resources and managing updates effectively. Despite the challenges, such as extensive data management and noise mitigation, guidance from our professors and self-initiated learning helped us navigate and gain a deeper understanding.

Moreover, differences in writing style were harmonized through collective discussions, achieving a unified voice in our output. Multitasking was our strength, allowing us to delve deeply into the project and fostering a broad and profound project understanding, which effectively helped us explore and develop eight unique refinements for our model and achieve favored results. Ultimately, this journey has enriched us both technically and in the art of teamwork.

2. Introduction

The beauty industry is currently experiencing a significant upsurge worldwide. According to Statista (2023), worldwide, this industry has generated revenue of £477 bn in 2023. Particularly in the United Kingdom (UK) this market generated a revenue of £13 billion in 2023. According to ReporterLinker (2021), the beauty industry is expected to grow by 4.82% by 2026. Relatively, according to (GlobalData, 2022) in the United Kingdom, it is expected to grow by 2% by 2026. This percentage may seem low; however, it signifies a steady and sustainable rise, reaffirming the industry's upward trajectory and its ability to continue drawing consumers' attention, spurring competition and driving economic growth.

Amidst the beauty industry’s growth, many competitors have emerged vying for a slice of the market. This has created a saturated environment due to which brands have recognized that differentiation is essential not only through conventional marketing and promotional strategies but also by strategically targeting niche markets. This approach goes beyond the surface and delves into the specific needs and preferences of diverse consumer groups. Consequently, brands are venturing into the realm of more intricate segmentation within product categories. This multifaceted segmentation not only allows them to carve out their unique positioning but also enables them to explore and exploit untapped opportunities within these niches, thereby maximising their market presence and appeal.

For instance, due to awareness of skincare routines with various steps, cleansing, toning, moisturising, and more, through social media platforms, consumers are informed about various skin concerns leading to demand for highly personalised products.

Brands are meeting this demand through different means. For instance, they are introducing products that can be incorporated in steps of the skincare routine which targets skin concerns such as wrinkles, acne, enlarged pores, redness and more. Moreover, they are introducing highly tailored products, for example, since pollution is on the rise, there is a niche market of consumers that have a concern of protecting their skin against pollutants that are present in the environment. Hence, Drunk Elephant's "D-Bronzi Anti-Pollution Sunshine Drops' ' provides a bronzing effect but also protects against pollution, contributing to a healthier and glowing complexion. Moreover, BareMineral’s “Complexion Rescue Tinted Hydrating Gel Cream” along with protecting against UV rays, also shields the skin from blue light that skin is exposed to due to screen time. As a result, such high demand of highly tailored products, has caused the product matrix to expand exponentially, posing a paradox of choice for the consumers.

Amid this, the demand for a personalized products are soaring. Research indicates that 71% of shoppers seek personalized interactions, with the absence of such experiences leading to disillusionment for more than 75% of customers (McKinsey and Company, 2021). According to McKinsey's study, personalization can have a direct bearing on customer loyalty and retention rates.

Moreover, advancements in technology and artificial intelligence are unlocking avenues for personalized experiences. This includes the use of chatbots for streamlined navigation, AI-powered makeup try-ons, AI driven shade matching, AI automating operations, AI self-driving cars, AI car manufacturing, AI hotel staff assistance, AI person-customer service, AI powered thermostats, AI recipe maker, AI powered video surveillance systems, etc. These technologies are not only enhancing user experience but also offering actionable insights to brands regarding consumer preferences.

Our collaborator, Yuty, is a nascent company established in 2020 in the UK, by Simi Lindgren. Yuty had an initial funding of £500k and has a goal of hitting £1 million in revenue by 2023. It aspires to harness these technological advancements to offer enhanced, personalized beauty solutions. Positioned as an online retail platform, Yuty operates on a direct-to-consumer (DTC) business model, focusing on providing personalized beauty experiences and products directly to individual consumers. Yuty encounters competition not only from established businesses but also from various online and offline players with diverse business models such as business to consumer (B2C) and direct to consumer (DTC). In addition to contending with well-known UK brands such as Look Fantastic, Cult Beauty, Boots, Sephora, and Superdrug, Yuty also competes with other brands that are targeting a niche market such as Space NK, which is a niche beauty retailer that focuses on offering a curated selection of luxury beauty products from various brands. This competitive environment underscores the challenge for Yuty to leverage technology effectively, differentiate itself, and establish a unique position in the market.

The central business problem that this project is focused on is to establish a machine learning model for Yuty that can accurately identify facial skin imperfections from user-uploaded selfies.

With this, Yuty aims to correct the consumer's self-perception regarding skin issues and to offer tailor-made product recommendations.

Recent advancements in artificial intelligence (AI), especially in visual image recognition, have been ground-breaking. Examples include its application in autonomous driving, facial recognition, natural language processing (NLP), etc. However, its application in skin assessment has been predominantly in the medical field, with models like U-Net being prominent for detecting ailments like acne and skin cancers. For example, Liu, Mou, Zhu, & Mandal (2020) introduced an enhanced U-net model for precise skin lesion segmentation, specifically targeting melanoma detection and utilise a transfer learning approach that combines U-Net and DCNN-SVM to enhance the accuracy of segmentation and detection in skin lesion images. Moreover, its commercial application, especially in the beauty sector, is still in its infancy, with only a few players making strides. For example, an article in SailThru (2020) highlighted L'Oreal, CoverGirl, Estee Lauder and Sephora as the few players that are utilising AI for maximising profit through personalisation.

2.1. Business Problem

The central business problem is to develop an automated and reliable system that can analyze customer selfies and detect various skin concerns with high accuracy. This system should highlight the detected concerns to provide a clear visual representation to the customers.
The following are the objectives of establishing such system:

- Accurate Skin Concern Detection: The deep learning model must accurately identify different skin concerns, ensuring that potential issues are not overlooked, and that false positives or negatives are minimized.
- Real-time Analysis: The system should process customer selfies in real-time, allowing for quick and seamless user experience. This is crucial for applications like virtual try-on, instant recommendations, or remote consultations.
- Generalizability: The model should be designed to work across diverse demographics and skin types, accommodating variations in lighting conditions, skin tones, and facial expressions.

2.2. Business Applications

Such a deep learning model will help businesses to improve their performance and position in this competitive market. There are several potential applications for such a skin concern detection model in the beauty and healthcare industries, which include:

- Personalized Skincare Recommendations: The model can detect various skin concerns, such as acne, dark spots, or redness, helping customers receive personalized skincare product recommendations that target their specific needs.
- Virtual Try-On for Skincare Products: The trending virtual try-on feature can be efficiently used by incorporating this model. By highlighting skin concerns in real-time, customers can virtually try different skincare products or makeup to see how they address their specific issues before making a purchase.
- Remote Dermatology Consultations: The model can assist dermatologists in remote consultations by pre-analysing customer selfies and flagging potential skin concerns, streamlining the diagnosis process.
- Beauty and Cosmetics Marketing: Companies can use the system to analyze customer selfies shared on social media to identify trends in skin concerns, helping them tailor marketing campaigns for specific demographics.
- Self-care and Skincare Tracking: Users can leverage the system to track the progress of their skincare routines over time by regularly uploading selfies, providing insights about the effectiveness of various products.

However, there are some challenges associated with this model. Customer selfies can vary widely in terms of image quality, lighting conditions, angles, facial expressions, and skin tones, making accurate and consistent segmentation challenging. Moreover, collecting a diverse and labelled dataset of customer selfies with ground truth annotations for skin concerns can be challenging and time-consuming. Furthermore, deep learning models, especially complex architectures, are often considered black boxes, it can be challenging to explain their predictions, which is critical for the successful implementation and adoption of the deep learning model in the beauty market.

Building a robust deep learning model for skin concern detection and highlighting from customer selfies has the potential to revolutionize the beauty and skincare industry by providing personalized recommendations and real-time analysis. By addressing business problems and objectives effectively, this system can offer enhanced customer experiences, improve skincare product sales, and facilitate remote dermatology consultations, among other practical applications. However, overcoming the challenges associated with data variability and interpretability is critical for the successful implementation and adoption of the system in the beauty market.

2.3. Research Questions

This project caters to the primary questions which arise about ways to design the model, improve its performance, and broaden the application boundary of the model technically. Although these questions primarily come up, we also explored to answer some of the more technical questions through this project:

I. Which deep learning architecture is best suited for a deep learning model aimed at detecting nine different skin concerns?
II. How does the performance of the deep learning model vary using data augmentation techniques?
III. How can transfer learning enhance the performance of this deep learning model?
IV. How does ensemble learning aid in improving the overall performance of this model?
V. How to deal with data imbalances to enhance model performance, focusing majorly on issues like imbalances across the 9 skin concern types, imbalances between skin concern areas and non-skin concern areas?
VI. How can issues related to sample size affect the performance of the model and how to solve it?

Building upon commitment and the challenges encountered, this research process has been nothing short of rigorous and enlightening. Eight experiment groups were tested, each armed with varying architectures, aiming to identify the most promising combination for Yuty's unique needs. Next, eight distinct enhancement techniques were fine-tuned according to the ResNet50 based U-Net model. Specifically, this model excels in detecting prominent skin concerns, especially Wrinkles Fine Lines, Redness, and Dark Circles, and has demonstrated a

14% improvement in IoU compared to its initial version. The further report includes the reasoning behind our chosen model and the solutions to our proposed research questions.

3. Literature Review

3.1. Image Segmentation

Image segmentation is a computer vision technique that involves dividing an image into meaningful and coherent regions or segments (Minaee, et al., 2022). The primary objective of image segmentation is to partition an image into multiple segments, each representing a distinct object or region of interest within the image.

The process of image segmentation aims to extract the boundaries or outlines of objects in an image, allowing for a more detailed understanding and analysis of the image content. By segmenting an image, it becomes possible to isolate and identify individual objects or regions, which can be useful for various applications such as object recognition, object tracking, scene understanding, and image editing (Haralick & Shapiro, 1985).

3.2. Types of Image Segmentation

Different types of image segmentations can be categorized based on the level of granularity or the desired output. In binary segmentation, the image is divided into two regions: foreground and background. Each pixel is labeled as either belonging to the foreground (object of interest) or the background (everything else) (Yang & Buenfeld, 2001).

Semantic segmentation takes a step further by assigning a semantic label to each pixel in the image, allowing for the identification of different object classes (e.g., person, car, tree). This approach aims to capture a higher level of understanding of the image content by associating meaningful labels with each pixel (Guo, Liu, Georgiou, & Lew, 2018).

Instance segmentation assigns semantic labels to pixels and distinguishes individual instances of objects. This enables the differentiation between multiple objects of the same class and assigns a unique label to each instance (Hafiz & Bhat, 2020).

In contrast, boundary segmentation's focus is on detecting and outlining the boundaries or edges of objects in an image. Its objective is to identify the precise boundaries between different objects or regions, which can prove beneficial for further analysis or object recognition tasks (Wang, Chen, Ji, Fan, & Li, 2022).

3.3. Techniques of Image segmentation

With a comprehensive understanding of the various types of image segmentation, the focus now shifts to the techniques employed to achieve these segmentation goals, shedding light on the techniques that underpin the extraction of meaningful regions from images.

These techniques as discussed by Szeliski (2010) and Felzenszwalb & Huttenlocher (2004), include thresholding, edge-based, region-based, clustering-based, and graph-based methods. Thresholding involves segregating pixels based on intensity relative to a set threshold. Edge-based techniques identify boundaries using intensity changes, while region-based methods group pixels by shared criteria. Clustering-based approaches employ algorithms like k-means, and graph-based methods use graph algorithms to partition images. These techniques collectively provide diverse avenues for achieving effective image segmentation.

3.4. Deep Learning in Image Segmentation

Deep learning is a subset of machine learning that utilizes artificial neural networks with multiple layers to learn and represent complex patterns in data (Atika, Nurmaini, Partan, & Sukandi, 2022). It has demonstrated remarkable success in various computer vision tasks, including image classification, object detection, and image segmentation.

Deep learning approaches, particularly convolutional neural networks (CNNs), have revolutionized image segmentation. Fully Convolutional Networks (FCNs), U-Net, and Mask R-CNN, as discussed in the following section, are popular architectures used for semantic and instance segmentation tasks. These models learn to directly predict pixel-level segmentation masks by leveraging large, labelled datasets.

These techniques are not mutually exclusive, and often a combination of methods is used to achieve accurate and robust image segmentation results. The choice of technique depends on the specific requirements of the application and the characteristics of the images being segmented.

3.4.1. FCNs

Fully Convolutional Networks (FCNs) represent a significant advancement in image segmentation. Departing from traditional CNNs, FCNs maintain spatial information by replacing fully connected layers with convolutional ones (Long, Shelhamer, & Darrell, 2015). This enables FCNs to generate pixel-wise predictions, vital for tasks like semantic and instance segmentation. Incorporating skip connections merges fine details and contextual information, enhancing accuracy. FCNs' adaptability to various input sizes makes them valuable across applications, from medical imaging to autonomy. FCNs redefine pixel-level analysis, transforming intricate visual data into meaningful interpretations.

3.4.2. Mask R-CNN

Mask R-CNN, proposed by He et al. (2017), is an extension of the Faster R-CNN object detection framework. It combines object detection and instance segmentation, allowing for the detection of multiple instances of objects and generating pixel-level masks for each instance. Mask R-CNN has been successfully applied in various applications, including medical imaging and scene understanding, where accurate instance segmentation is essential.

3.4.3. U-Net

U-Net, introduced by Ronneberger, Fischer & Brox et al., (2015), is a popular architecture for image segmentation. It consists of a symmetric encoder-decoder network with skip connections that preserve spatial information during up sampling. The skip connections concatenate feature maps from the encoder to the corresponding decoder layers, aiding in the precise localization of objects. U-Net is known for its ability to handle limited data and small object structures effectively.

Moreover, U-Net is recognized for its ability to address the challenges posed by medical images, which often feature intricate structures, variations in appearance, and complex backgrounds (Litjens, et al., 2017). U-Net is utilised for its design, as encompassing a contracting path for feature extraction and an expansive path for detailed localization, is highlighted as a suitable approach for handling medical imaging challenges.

Du et al. (2020) elaborate on how the U-Net architecture finds extensive application within medical image analysis, particularly for tasks such as lesion segmentation, anatomical segmentation, and classification. The widespread adoption of U-Net in this domain stems from its inherent ability to precisely segment critical features of interest, proficiently handle and impartially assess medical images, ultimately leading to enhanced diagnostic accuracy in medical imaging.

Findings of Roth, et al., (2018) acknowledge the significance of deep learning techniques in advancing medical image analysis and delve into specific architectural choices, with a notable focus on U-Net. The U-Net architecture is described as effective in accurately segmenting anatomical structures and identifying pathological lesions within medical images. The article highlights the unique design of U-Net, which features a contracting path for feature extraction and an expansive path for precise localization, is noted as a key factor contributing to its success in these tasks.

- *ResNet-50*

ResNet-50 is a specific variant of the ResNet architecture, comprising 50 layers. This version strikes a balance between complexity and computational efficiency, making it well-suited for various applications, including image segmentation. With ResNet-50, the architecture's depth allows it to capture intricate features, while the skip connections ensure information flow is smooth, benefiting both feature extraction and context preservation (Renjun, Junliang, Yi, & Mengcheng, 2022).

Furthermore, ResNet-50's pre-trained weights enhance its proficiency, offering an advantage in tasks like transfer learning particularly when labelled data is scarce. Employing ResNet-50 as an encoder in the U-Net architecture leverages these advantages, leading to more accurate and effective segmentation results.

As such, ResNet-50 emerges as a strategic choice for enhancing U-Net's performance in image segmentation tasks, especially when fine-tuning with specific datasets is a priority. The interplay between U-Net and ResNet-50 amplifies their individual strengths, culminating in a potent combination for robust image analysis.

- *ResNet-50 compared with other encoders.*

When considering the selection of an encoder for U-Net, it's essential to examine how ResNet-50 compares with other commonly used encoders.

VGG, although capable of capturing basic features, may demand more parameters to learn complex features compared to ResNet-50. Inception's multiscale feature extraction through parallel convolutions is advantageous, but ResNet-50's balance between complexity and performance often aligns better with segmentation tasks (Atliha & Šešok, 2020). EfficientNet's parameter efficiency and adaptability through scaling coefficients make it versatile, but ResNet-50's deep structure enhances its capacity to capture intricate patterns (Abedalla, Abdullah, Al-Ayyoub, & Benkhelifa, 2021). DenseNet's feature reuse is valuable, yet the associated memory usage might not always align with segmentation requirements (Zhang, et al., 2021).

ResNet-50 emerges as a strategic choice for enhancing U-Net's performance in image segmentation tasks, particularly due to its depth, skip connections, and transfer learning capabilities. While other encoders offer unique strengths, ResNet-50's ability to harmonize depth, context preservation, and pre-trained weights within the U-Net framework proves advantageous for precise and robust segmentation. As a result, ResNet-50 stands as a compelling encoder option that synergizes effectively with U-Net's architecture, amplifying its efficacy in diverse image analysis scenarios.

3.5. Advancements in Deep Learning for Medical and Skin-Related Applications

The following is an exploration of advancements made in deep learning for applications related to medical, dermatological and beauty industries.

3.5.1. Deep learning advancement in the Medical Industry

In recent years, deep learning-based segmentation has shown significant progress in medical imaging. Studies by Hesamian et al. (2019) and Salehi et al. (2017) discuss the application of U-Net for the segmentation of organs and lesions in medical images. These works demonstrate the effectiveness of deep learning in handling complex medical datasets and achieving accurate segmentation results.

Medical Image Segmentation using 3D CNN**s:** Salehi et al. (2017) proposed using 3D CNNs for medical image segmentation, specifically for MRI and CT scans. 3D CNNs capture spatial relationships along the third dimension, enabling accurate segmentations of volumetric medical data.

3.5.2. Deep Learning for Dermatological Image Analysis

- *Skin Lesion Segmentation*

Deep learning has also been applied to skin lesion segmentation, which is essential for diagnosing and monitoring skin conditions. Research by Esteva et al. (2017) showcases the use of deep learning for melanoma detection, a deadly form of skin cancer. The study employs a CNN-based model for precise segmentation of skin lesions, aiding in the early detection and diagnosis of melanoma.

Haenssle et al. (2018) demonstrated the use of deep learning for skin cancer detection. Their study employed CNNs to classify skin lesions in dermoscopic images, achieving performance comparable to dermatologists. This breakthrough has paved the way for using deep learning in skin-related applications.

- *Attention Mechanisms in Segmentation*

Hasson et al. (2019) introduced an attention-based mechanism for medical image segmentation. The attention mechanism enhances the relevant features while suppressing irrelevant regions, improving segmentation accuracy. This approach has shown promising results in segmenting medical structures with complex shapes.

These examples collectively illuminate the far-reaching impact of deep learning in reshaping the landscape of medical image analysis, delivering enhanced accuracy and insights across diverse medical domains.

3.5.3. Deep learning advancement in the Beauty industry

Recent advancements in deep learning-based segmentation have demonstrated promising results in beauty-related applications, providing opportunities for improved diagnosis and virtual try-on experiences.

- *Diagnosis*

In the beauty industry, image segmentation is widely used for tracking acne, with studies showing advancements in methods and models. Kittigul & Uyyanonvara (2016) developed an automatic acne detection system using Haar Cascade for face extraction, GrabCut for background removal, and Heat mapping for visualization. The Adaptive Thresholding and BLOB detection marked acne areas. Zhang et al. (2022) research highlights theory-driven image processing and data-driven DL methods for acne, favouring TransUNet for accuracy, recall, precision, and F1-Score.

- Virtual try-on experiences

Study by Li et al. (2019) explore the use of Mask R-CNN for virtual try-on and makeup recommendation. These applications allow customers to visualize how skincare products and cosmetics will appear on their faces, considering specific skin concerns identified through segmentation.

3.6. Image Segmentation for Skin Concern Detection

Image segmentation has become a crucial part of personalized skincare and the beauty industry. This approach helps identify various skin concerns accurately, allowing for tailored solutions. In this exploration, we'll dive into the significance of image segmentation for detecting skin concerns.

The skincare industry has experienced significant growth lately due to changing consumer preferences, technology advancements, and evolving beauty perceptions. According to a report published by Fortune Business Insights (2022), focused on forecasting skincare industry trends, highlighted that increased awareness of personal care among all age groups has driven demand of skincare and this is causing a lifestyle shift to focus on reducing skin concerns such as wrinkles, dark spots, acne etc. Furthermore, as Grand View Research (2020) has described that the demand for personalized skincare products has increased across various age groups, as consumers recognize the need for customized daily skincare routines due to different skin types, concerns, tones, preferred ingredients, and lifestyles. Due to this rise in demand research has been detected to image segmentation to detect skin concerns.

Ko & Cheoi (2021), present a technique for detecting and segmenting facial imperfections using image processing. The proposed method aims to identify various types of skin abnormalities simultaneously from facial photographs taken with a standard smartphone camera. The technique involves adjusting illumination changes using an illumination removal algorithm and employing a flexible Gabor filter for adaptive detection of skin disease objects. The authors also introduce a grouping technique for identifying and categorizing abnormal objects and regions. They address the challenges of detecting scattered and distributed skin diseases by using the density-based spatial clustering of applications with noise (DBSCAN) algorithm for grouping. The proposed approach offers a comprehensive method for detecting and segmenting facial imperfections, contributing to improved understanding and analysis of skin conditions.

An article by Yoon, Kim, Lee, & Yoo (2023) focuses on facial skin analysis using deep learning to simultaneously segment wrinkles and pores. The study proposes a novel approach using the

U-Net model with attention schemes to enhance important areas. The authors highlight the significance of morphological structures over color-based analysis.

Work of Zheng et al. (2022) focuses on the automatic assessment and understanding of facial skin condition for various applications, such as health detection, skincare recommendations, and more. It addresses challenges in collecting diverse selfie data and aims to ensure accurate assessment by detecting different skin features. The method utilizes a two-phase annotation scheme, involving dermatologist guidance to train volunteers for annotation. The Unet++ network architecture is employed for feature detection. The study demonstrates that this approach robustly detects acne, pigmentation, and wrinkles across diverse factors like ethnicities, skin tones, severity levels, age groups, and lighting conditions in selfie images.
Skin segmentation is crucial for personalized skincare, involving isolating skin regions from images. Color-based methods are common.

A paper by Li, et al. (2020) identifies the importance of accurate identification of a person's skin quality and appropriate product usage are essential to avoid potential skin damage. The study delves into the application of machine learning and deep learning algorithms to create an intelligent recommendation platform for facial skin care. The research employs YOLOv4, an innovative object recognition algorithm, to detect crucial features in facial images and segment them into sub-images of regions of interest (ROI). These sub-images are then analyzed by a multi-label model to identify and evaluate problematic areas using the YOLOv4 identifier. Additionally, image processing techniques are employed to preprocess, enhance, and extract features from sub-images, forming feature vectors used to train the multi-label classification model for skin condition classification.

3.7. Conclusion

Image segmentation is reshaping the landscape of personalized skincare and the beauty industry. This technique enables the accurate identification and isolation of various skin concerns, paving the way for personalised and effective solutions. The skincare industry's growth, driven by evolving consumer preferences and technological advancements, has led to a surge in demand for personalized skincare products that cater to diverse needs and preferences.

The significance of image segmentation is highlighted through its application in medical and dermatological contexts in current times. Research endeavours in the medical field, as well as the beauty sector, underscore the transformative potential of this technique. The ability to precisely segment skin regions and identify concerns like acne, wrinkles, and pigmentation demonstrates its efficacy in enhancing both diagnostic accuracy and cosmetic understanding. Moreover, the U-Net architecture emerges as a powerful tool, adept at handling various challenges like class imbalance, data augmentation, and transfer learning. Its adaptability holds the potential to reshape skincare diagnostics, offering precise and individualized care.

Furthermore, businesses in the skincare and beauty industry can leverage the insights gained from image segmentation research to revolutionize their products and services. By integrating advanced techniques like deep learning and image analysis, companies can offer personalized skincare solutions that precisely address individual concerns. This technology allows for the accurate identification of specific skin issues, such as acne, wrinkles, and pigmentation, enabling brands to develop targeted products and tailored regimens. Moreover, businesses can enhance customer experiences by providing virtual try-on experiences, allowing customers to visualize the effects of skincare products before making a purchase. Such innovations not only improve customer satisfaction but also establish brands as leaders in offering data-driven, effective, and personalized skincare solutions in an increasingly competitive market.

In sum, image segmentation emerges as a dynamic and integral tool in the personalized skincare domain. Its contribution in offering tailored solutions, coupled with advancements in deep learning and technology, shapes a promising future for skincare diagnostics and personalized treatment recommendations. This ever-evolving field holds the potential to revolutionize how we understand and address skin concerns, ultimately leading to improved well-being and self-care practices.

3.8. Research Gap

This study aims to fill a gap in the field by developing a deep learning model that can accurately detect various skin concerns based on users' selfies. To achieve this, we're using a UNET architecture with ResNet 50 as an encoder. This research holds significance because it's tailored to meet the needs of the beauty industry. By doing so, our study adds to the existing body of knowledge while aligning with the unique requirements of the beauty market. In essence, we're contributing valuable insights that can enhance skincare diagnostics and personalized solutions. By improving skin problem detection and analysis, our research can lead to enhanced customer experiences, more effective product recommendations, and improved overall satisfaction in the beauty industry.

4. Research Methodology

4.1. Dataset

4.1.1. Data Collection

Our dataset comprises a total of 22,132 images and a corresponding JSON file. These resources have been generously provided by Yuty.

The images encompass a diverse range of human faces, featuring individuals of varying genders, ages, and ethnicities. This diversity introduces a spectrum of skin tone and skin conditions. Additionally, some subjects are depicted wearing accessories like sunglasses, glasses, hats, ribbons, and other facial coverings, posing a challenge to accurately identify the precise facial region. Moreover, while most images spotlight a solitary human face, a subset portrays multiple faces, and a few exhibits distorted facial features due to technical nuances in the image collection process. Addressing these anomalies will be part of our subsequent data cleaning efforts.

In the original JSON file, each file name corresponds to a data item, and each data item contains a "filename" key and a "regions" key. The value of the "regions" key is a list of regions, each containing the keys "shape_attributes" and "region_attributes". The X-coordinates and Y-coordinates are connected to a block of polygonal areas, which means skin concern areas of various classes. Each image file can have multiple skin concern regions, with different shape attributes and region attributes for describing a specific region in the image and the results of skin analysis in that region.

Our focus for this project is to find and understand nine specific skin concerns that Yuty has presented. These problems include skin concerns like Redness, Wrinkles and Fine Lines, Uneven Skin Tone, Dark Circles, Clogged Pores, Dark Marks and Hyperpigmentation, Blemishes, Acne, and Freckles and Beauty Spots.

4.1.2. Data Cleaning

To reduce the effect of noise caused by the provided data, we first manually screened a dataset of 22,132 images, before applying data quality screening to the remaining 14,515 images.

- *Manual Screening*

It is found that a significant number of the 22,132 images contain blocked or covered human faces. To remove those images without a complete and correct human face, certain criteria have been established in the manual screening process. The first rule is to remove images in which the face is covered, partially covered, the eyes are covered, glasses or sunglasses are worn, too much make-up is applied or tattoos are visible. The second rule was to remove images of people with their hair covering their eyes, which is most often the case in images of Asian girls. The third rule required the removal of images in which the individual's face was not directed straight ahead. For instance, faces that were angled left or right or up or down should be excluded manually. For the sake of the model's performance, the final rule was about the removal of images with multiple faces, which could have a detrimental effect on our results.

Prior to manual screening, there were 22,132 images and after manual screening, there were 14,515 images left, giving an average manual screening rate of about 65.6%. Meanwhile, we need to get the JSON file corresponding to the 14,515 images ready. First, we read the list of filenames for the manually screened images, then we load the images in google drive catalogue based on those filenames and convert each image into a 3-channel Numpy array.

- *Data Quality Filtering*

Once we have the manually screened images ready and their corresponding JSON files prepared, we start filtering the data according to image quality, which involves eliminating images that are of lower quality than the remaining images. As is common practice, we choose three quality filtering measurements: the mean pixel value, the variance and the noise variance of each image.

An obvious relationship exists between the three measurements chosen and image quality. Specifically, the average pixel value reflects the brightness of the image. A higher average pixel value usually indicates a brighter image, while a lower average pixel value indicates a darker image. In image quality assessment, the average pixel value can be used to detect whether an image is overexposed or too dark. As for variance, the variance of an image indicates the degree of dispersion or range of variation of pixel values in an image. It measures the extent to which the pixel values in an image deviate from the mean value of the image, with a higher variance indicating a greater fluctuation in the pixel values in the image. Images with higher variance may have richer detail and contrast. As for noise variance, it is a measure of the strength of random noise present in an image. Noise is the fluctuation of pixel values due to random disturbances introduced during image acquisition, transmission or processing. High noise variance implies poor image quality because noise reduces the sharpness, detail and contrast of an image.

Gradient and sharpness, which are more related with the edge information, are not used as measurements because the faces in given images are in the middle in most cases, while the edge information has little to do with the facial region.

Next, we set up three percentile thresholds for them as quality filtering criteria. Compared with using average value or median value as threshold, using percentile value as threshold has an advantage in the following two aspects. Firstly, percentile threshold filtering can mitigate the effects of outliers to a certain extent because it focuses on a portion of the overall distribution rather than relying on statistics such as the mean or variance, which are susceptible to the interference of outliers. Moreover, percentile threshold can be adjusted to fit different data characteristics by adapting it to the distribution of the actual dataset. We can choose different percentiles according to specific needs to meet the filtering requirements.

In the practice, images with average pixel values between 10~90% are retained to remove overexposed or too dark images. 90% of images with higher variance are retained because images with higher variance have richer contrast. 90% of images with less noise variance are retained because less noisy images are of higher quality in most cases. We can also adjust these thresholds if necessary. Images that survive though quality filtering are labelled as "filtered images". For example, if we start with 500 images, we will acquire 265 images after manual screening and 175 images after filtering.

4.2. Evaluation Metrics

This model employs Intersection over Union (IoU), Recall, and Dice Coefficient as performance evaluation metrics. Among these, IoU provides a clear measure of the alignment between predicted and actual regions, and it is widely used in tasks like object detection and image segmentation. Recall measures the proportion of actual positive instances correctly identified by the model. Like IoU, Dice Coefficient quantifies the similarity between predicted and actual regions.

The reason for not using the F1 score in this model is twofold: Firstly, the Dice Coefficient is essentially a variant of the F1 score. Secondly, although the F1 score is commonly utilized, it might not perform well when dealing with highly imbalanced datasets, as it can be biased by the majority class and our dataset exhibits certain imbalance characteristics, which will be elaborated further.

4.3. Data Pre-processing

To start with, we split the dataset of filtered images into the training set (80%), the validation set (10%) and the test set (10%), and set a random seed to maintain the consistency of the random split. The image arrays are split into multiple chunks. Such processing can be used to input large size images in a sliding window fashion, rather than using the whole image directly for training.

Each sliding window is set to be a small block of 256x256 pixels, where the height is 256 pixels, the width is 256 pixels and the number of color channels is 3. This means that the original large image is split into small blocks, each of which is an RGB image of 256x256 pixels. The sliding window step is set to 256 pixels while keeping the 3 color channels constant. We chose a window shape of 256 x 256 because it is a relatively common choice and can balance the capture of details and context information, which is suitable for many segmentation tasks (Zhang, et al., 2020; Meena, et al., 2022). Additionally, we chose a step size of 256 because it simplifies assembling the original image at the end.

Afterwards, we start creating the mask. Firstly, a mask for each image is created based on the data in the JSON file, each mask represents the labelling information for different regions in the image. Such processing can assign different labels to each region in the image, which can be used for training and evaluating the image segmentation model. The loop in the code will iterate through all the data and then create the corresponding mask for each image. When iterating over data with missing images, the corresponding masks will not be created.

The mask list is then split into the training set, the validation set and the test set. The masks are split into multiple chunks as before. Such processing can be used for input in a sliding window fashion for large sized mask images, rather than directly using the whole mask image for training. After performing the sliding window segmentation on the mask images, each image is split into multiple chunks of 256 x 256 pixel size and each chunk contains 9 channels for representing different attribute information. The sliding window shape and sliding window step size were set to 256 pixels x 256 pixels x 9 channels.

Before modelling, we need to prepare the data by dividing the image data and the corresponding mask data into the training set, the validation set and the test set and pre-processing them accordingly. As it is an image segmentation task, we need to assign a label to each pixel (which category it belongs to), so the mask is shaped as (height, width, number of channels), where each pixel has multiple channels of the mask, and each channel represents a different label.

Next, by traversing the masks in the training set, we need to find the indexes of the samples for which the pixel values of all the channels are zero. These samples have no target object for the image segmentation task and can be considered as invalid samples. Then remove all the samples in the training set whose masks are all background so training set contains only samples that have a target object. Finally, the image data should be normalized by scaling the pixel values from the integer range [0, 255] to the floating-point range [0, 1]. Normalization helps in the training and optimization process of the model by keeping the data in a relatively uniform range.

4.4. Model selection

Our objective is to analyse facial skin in selfies, determine the presence of nine different skin issues, and perform area annotation and problem recognition for the same. For this project, we have opted for the U-Net convolutional neural network, which was specifically designed for medical image segmentation and has shown notable performance in the field of biomedical image recognition. Based on the U-Net architecture, we've crafted three models: U-Net, U-Net++, and U-Net with 6 different encoders. Ultimately, we will select the optimal model based on its performance on the training and validation datasets.

The following three sections will delve into the reasons behind selecting U-Net as the principal architecture, why these three models were chosen within the U-Net framework, and for the third model's encoder, the reasons for choosing VGG16, Xception, ResNet50, ResNet101, EfficientNetB4, and DenseNet201 as alternative encoders, respectively. Finally, we will describe the steps and details involved in the model's actual implementation.

4.4.1. Rationale for choosing the U-net as core architecture.

When performing image segmentation tasks, especially in medical or other scenarios where accurate and detailed segmentation is required, choosing a suitable model is crucial. In our project, we chose U-Net as the core model, and the following are our main considerations and rationale:

- *Origin and design goals of U-Net*

The application of deep learning to medical images has been growing rapidly from 2015 to 2016 and has become the dominating choice for medical imaging interpretation, while the vast majority of deep learning detection systems use convolutional neural networks (CNNs) to perform pixel classification (Litjens, et al., 2017). U-Net stands as a pivotal framework in CNN-driven semantic segmentation, frequently leveraged in medical imaging to segment lesions, delineate anatomical structures, and classify anomalies (Du, et al., 2020).

The detection task of image segmentation consists of finding the location and identifying the problem (Litjens, et al., 2017). Du et al. (2020) pointed out that 'the advantages of the U-Net network framework is that it not only accurately segments the desired feature targets, efficiently processes and objectively evaluates medical images, but also helps to improve the accuracy of medical image diagnose'. Furthermore, the U-Net has achieved decent results in a variety of biomedical segmentation applications (Ronneberger, Fischer & Brox, 2015).

Our project is to diagnose the skin concerns in the images, find the locations corresponding to the problem, and classify and identify nine skin problems, and the outstanding performance of the U-Net model in the field of medical image segmentation is one of the most important factors for us to choose this model.

- *Core Advantages of U-Net vs Others*

Despite the existence of advanced and structurally complex models such as Mask R-CNN, YOLO, etc., our decision to choose U-Net is based on the core strengths it demonstrates in the following areas:

I. Data efficiency

A significant advantage of U-Net is its performance on small datasets (Ronneberger, Fischer & Brox, 2015). With built-in data augmentation techniques, U-Net is able to make the most of limited labelled data to generate reliable and accurate segmentation.

In this project, our computing environment is Google Colab, and we do not have access to higher-end computing resources such as computing clusters. Due to the constraints of hardware performance, the images and their annotated data that we can input and process have their limitations. Therefore, the high efficiency and computational economy of U-Net especially meets our practical needs.

II. Computational efficiency

Although U-Net is designed for high accuracy, its structure is relatively simple. This means that relative to more complex models, U-Net has a much faster speedup, requiring only a relatively short training and inference time (Ronneberger, Fischer & Brox, 2015). While Mask R-CNN is a very powerful but also very complex model that requires significant computational resources and training time.

For our project, it is especially critical in contexts where we need to cater for users uploading images in real-time and giving real-time feedback, as well as limited computational resources.

III. Simplicity and Interpretability

U-Net has a relatively simple structure, which means it is easier to implement, debug and interpret. For our project, it is still relatively in the early stages of the entire project. So these features, facilitate rapid iteration and optimisation afterwards.

IV. Generality vs Specialisation

While models such as YOLO and Mask R-CNN perform well on target detection and instance segmentation tasks, U-Net is better suited to texture-rich, less varied image segmentation tasks. In particular, it may perform well on skin problem image segmentation tasks.

V. Adaptability and Scalability of U-Net

The success of U-Net is partly due to its special structure, which consists of a contraction path, which can be considered as an encoder, and an expansion path, which can be considered as a decoder. This structure means that the encoder can be easily replaced by other pre-trained models.

The use of pre-trained encoders means that U-Net can utilise features already learnt on large-scale datasets such as ImageNet and apply some of the features that are generic in nature to similar problems through transfer learning (Dey, 2018). This approach can help the model converge faster on a specific medical image dataset. Models with different encoders (e.g., VGG16, ResNet50) have different structures and properties. This means that they may have varying performance when dealing with different types of data or data of varying complexity.

Compared to complex models such as Mask R-CNN and YOLO, whose complex structures may make transfer learning less advantageous, although it is also possible to use pre-trained models, U-Net is much easier to combine with different encoders, providing great flexibility.

In conclusion U-Net presents us with a structurally simple and exceptionally efficient alternative. Although models such as Mask R-CNN are more versatile in the field of object detection, in our scenario we focus more on fine segmentation of the image rather than multi-object detection. In contrast, models such as YOLO focus on optimising the speed of real-time detection and may compromise on segmentation accuracy. U-Net finds the right balance between computational efficiency and segmentation accuracy, making it an ideal choice for our project.

In summary, given the unique requirements and certain constraints of our project, U-Net proved to be an appropriate choice. Its simplicity, efficiency and adaptability make it the model of choice for image segmentation tasks when the computational resources are limited.

4.4.2. Rationale for Choosing U-Net, U-Net++, and U-Net with 6 Types of Encoders

Among the myriad models based on the U-Net architecture, each possesses its own unique features and application scenarios. In this project, we've specifically selected three representative models for an in-depth exploration. Each of these models exemplifies the immense potential and versatility of U-Net in the realm of image segmentation. In the sections that follow, we will delve into the specific reasons behind choosing these three models and highlight their advantages and characteristics in practical applications.

U-Net: As the original architecture, it serves as a baseline for comparing with other models. Its simplicity and efficiency have made it widely favoured across numerous image segmentation tasks.

U-Net++: This is an enhanced version of U-Net, introducing additional skip connections and a nested structure to boost the model's performance. This implies it can capture more features and performs better in more intricate tasks.

U-Net with 6 Types of Encoders: Integrating U-Net with other pre-trained models allows us to benefit from the training of other models on large datasets. This approach allows us to capture more features and enhance the model's performance.

4.4.3 Rationale for Choosing VGG16, Xception, ResNet50, ResNet101, EfficientNetB4, and DenseNet201 as Alternative Encoders.

I. Diversity of Models

These 6 encoders bring to the table varied structures and depths, each with unique capabilities for feature capture. Having these alternative encoders aids in experimentally determining which structure is most suitable for a particular task.

VGG16 is relatively straightforward and not very deep but contains a vast number of parameters. It can act as a standard to gauge how simpler structures tackle the task.

ResNet50 and ResNet101, as their names indicate, comprise 50 and 101 layers respectively. This presents an ideal scenario to delve into the correlation between network depth and its performance. However, they might demand more computational resources.

Xception is an extreme convolutional network sculpted with the aim to utilize parameters effectively. One drawback is that its performance might not be optimal for certain specific tasks.

EfficientNet series is lauded for delivering commendable performance with comparatively fewer parameters. These structures were derived through automated searches for the optimal network structure. Yet, a downside is its high demands regarding training configuration, necessitating meticulous hyperparameter tuning.

DenseNet201 realizes feature reuse via dense connections, augmenting network efficiency. The number of parameters it requires is relatively less. However, a potential downside is a slower training process.

II. Advantages of Transfer Learning

VGG16, Xception, ResNet50，ResNet101，EfficientNetB4，DenseNet201 are all well-known pre-trained models in the field of deep learning. These models have all been pre-trained on ImageNet, which means that they have learnt a large number of general features, which can accelerate the training of the models on our dataset.

In summary, by choosing VGG16, Xception, ResNet50，ResNet101，EfficientNetB4，DenseNet201 as alternative encoders, it was possible to explore how different network architectures affect the model's performance, speed, and generalisation ability, in order to find the optimal encoder for this project.

4.4.4. Experimental Design

In this section, we first provide a brief description of the experimental design and training process, and then explain the rationale behind the choice of the loss function &the key parameters. Follows by the rationale of choosing the pre-trained face detection model, as well as the main libraries used.

- *Experimental Design & Training Process*

Combined with the previous data cleaning and preparation, the overall experimental flow can be divided into 4 main stages, as shown in **Figure *1***.

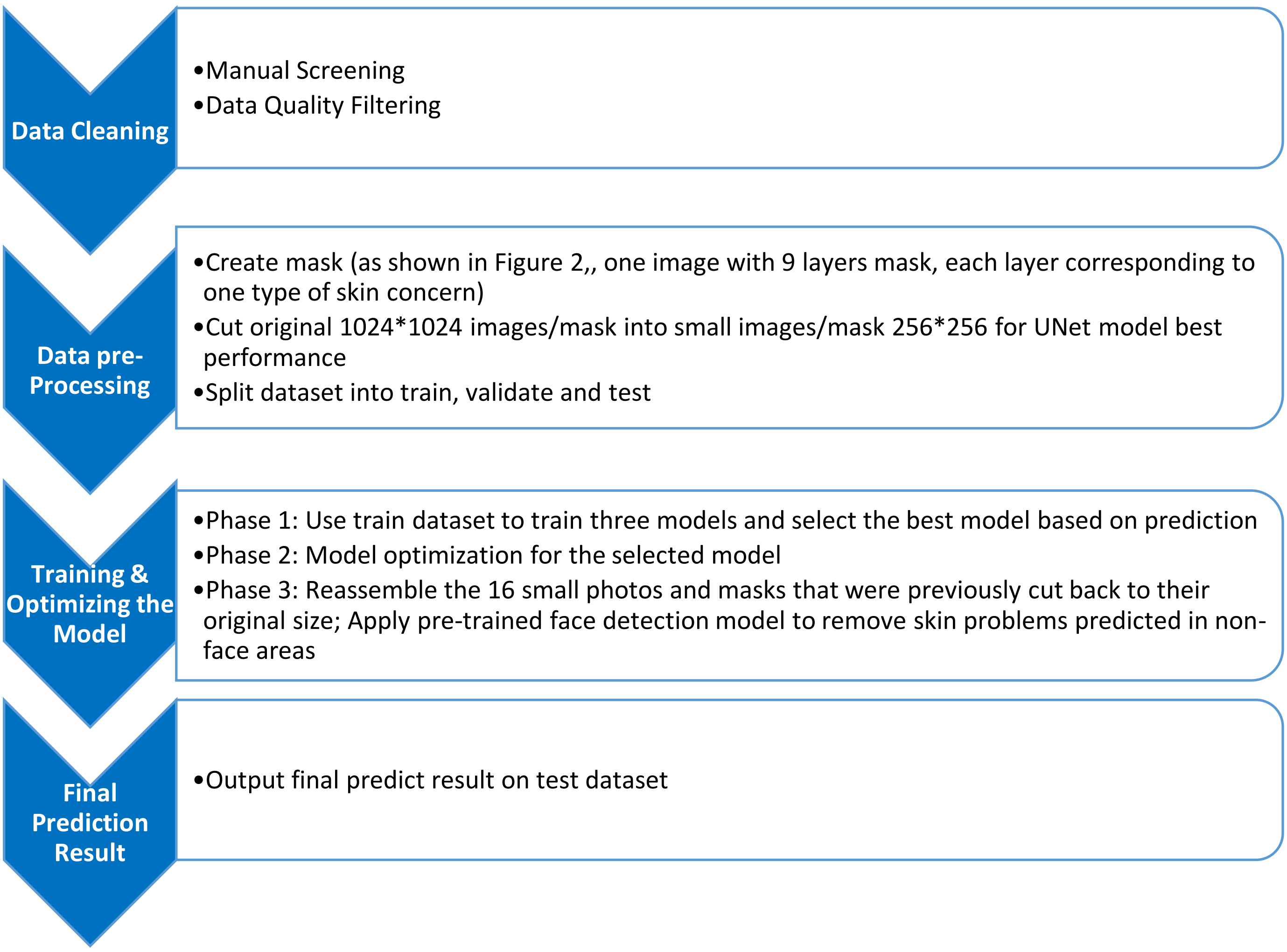


Figure 1: Experimental Flow

During the second stage ‘Data Pre-processing’, we transformed JSON file annotation data into a structured mask format that aligns well with the demands of semantic segmentation tasks. Each of these masks illuminates regions within an image, each associated with distinct attributes or classes. A single image comprises a mask with nine layers, where each layer is tantamount to a specific skin concern, as illustrated in **Figure *2***

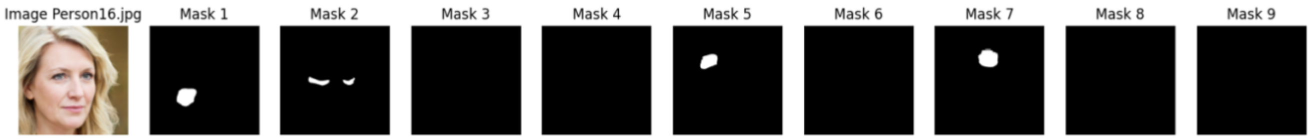


Figure 2: Example of selfie image with the 9-layer skin concerns masks (the sequence of mask: 'Redness', 'Wrinkles_Fine_Lines', 'Dark_Circles', 'Uneven_Skintone', 'Clogged_Pores', 'Hyperpigmentation_Dark_Marks', 'Blemish', 'Acne', 'Freckle_Beauty_Spot')

After this, we partitioned our dataset, cropping the original 1024x1024 images and their associated masks into smaller 256x256 windows using a stride of 256 to optimize U-Net model performance. Following data normalization, these images were designated as our 'X' variables, while their corresponding masks—indicating the presence and location of skin issues—were treated as 'Y' variables for the model. Of these, 80% were allocated to the training set, 10% to the validation set, and 10% to the final test set. Both the training and validation sets were then advanced to the third stage

The third stage 'Training & Optimizing the Model' can be dissected into three distinct steps. Firstly, we use the training dataset to train three Models (U-Net, U-Net++, U-Net architecture +various pre-trained encoder) and select the best model based on the prediction result. Among the pre-trained encoders we included 6 choices, resulting in 8 models and variants groups, as shown in **Table *1***, and we will choose the best one to enter the next stage of optimisation based on the performance in the training dataset and validation dataset.

Table 1: Details of 8 experimental groups

	Group 1	Group 2	Group 3	Group 4	Group 5	Group 6	Group 7	Group 8
Models	U-Net	U-Net ++	U-Net architecture+ pre-trained encoders					
Encoders	-	-	VGG 16	Xception	ResNet50	ResNet101	EfficientNetB4	DenseNet 201
Measurement Metrics	Select best model based on learning and prediction ability: Intersection over Union (IoU), Recall, and Dice Coefficient on training and validation dataset							

The second step focuses on model optimization. This encompasses refining the model's structure, addressing overfitting or underfitting issues, enhancing data quality and diversity, tuning hyperparameters, and mitigating data imbalance challenges.

In the third step, the 16 smaller 256x256 images and masks, which were previously segmented, are reassembled to their original 1024x1024 dimensions. Subsequently, predictions of skin issues in non-face regions are filtered out by deploying a pre-trained face detection model. This ensures the accuracy and relevance of the skin problem detection in the final output of the selfie.

- *The main function and parameter choices*

The main function and parameter choices of the model are default values or mainstream choices based on experience, these initial choices provide us with a starting point and will be adjusted appropriately according to the actual performance of the model in the second phase of the experiment.
Loss function: focal loss was chosen for this project mainly because of the possible imbalance between the 9 skin problems and the possible imbalance between the pixel area of the skin problem and the non-skin problem area, these two imbalances will affect the model prediction ability, and the focal loss function can be adjusted in the second phase of the experiment to solve these problems to a certain extent and optimise the results.
Learning rate: 0.0001 is a common choice that provides a steady training rate without causing the model to converge or diverge too quickly.
Optimiser: Adam is an adaptive learning rate optimiser that typically performs well in deep learning tasks.
Number of Epoch & Batch size: The model was trained using 100 iterations and a batch size of 32 was chosen.

4.4.5. The Rationale for Choosing MTCNN as Pre-trained face detection model

Regarding the selection of pre-trained face detection model, based on the maturity stability, real-time performance, accuracy, and robustness of the model, we chose OpenCV and MTCNN (Multi-task Cascaded Convolutional Networks) as the alternatives for face detection.

From the effect of face recognition presented on the image, as demonstrated in **Figure 3**, MTCNN recognises faces more accurately, with a better fit between the face detection frame and the face in the real photo. In addition, when we randomly sampled 200 photos and performed face recognition on them, and cross-checked the recognised face regions with the skin problems labelled in the original data, the result in **Table 2** showed that MTCNN had a higher IoU, i.e., a higher degree of overlap, which means its MTCNN has higher accuracy in terms of face detection frame. Combining these two reasons, MTCNN was selected for face recognition detection in this project.

Table 2: Average IoU between skin concerns and face detection frame

	Average IoU between face detection frame with skin concerns
OpenCV model	0.14
MTCNN model	0.17

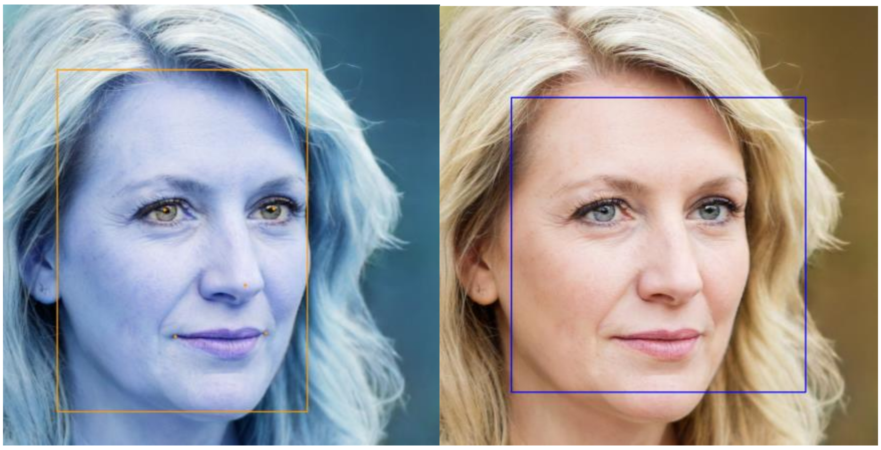


Figure 3: Face detection frame from MTCNN and OpenCV on image Person16

- *Main library utilised*

In this project, we mainly rely on the Keras library to build and train the model. keras provides a concise and efficient API that can easily build complex neural network structures. Meanwhile, the underlying use of the TensorFlow framework ensures computational efficiency and scalability.

With the above design and choices, we aim to build an efficient, robust, and accurate region segmentation model for skin problems.

5. Results and Analysis

5.1. Model Performance

In this stage, we randomly selected 327 samples post-data cleaning, and allowed 8 different groups to train on the training dataset and subsequently make predictions on both the training and validation datasets.

From Figure 5, it is evident that the foundational U-Net model and its upgraded counterpart, U-Net++, exhibited subpar performances in detecting skin anomalies. Their respective average IoU scores on the training dataset were a mere 12% and 2%. In the validation dataset, these figures were further diminished to 8% and 2%. A plausible rationale for this is the inherent simplicity of these models, potentially inadequate for capturing the intricate features associated with skin anomalies. Additionally, the samples might possess considerable diversity and imbalance, making convergence during training more elusive for these models.

In contrast, the rest six model groups, integrating the U-Net architecture with various pre-trained encoders, displayed markedly superior performances. Their average IoU scores on the training dataset hovered around 80%, and in the validation dataset, these figures reached about 20%. This underscores the potency of introducing pre-trained encoders, which, having undergone extensive preliminary training, have already assimilated knowledge regarding image characteristics. This foundational knowledge evidently renders them more adept for our specialized task.

However, a glaring concern for these six models is their pronounced susceptibility to overfitting. While these encoder groups performed admirably on the training dataset, their prowess diminished notably on the validation dataset. This indicates a potential over-attunement to the training dataset, making them less adept at handling novel, unseen data. The intricate nature of these encoders might lead them to overly fixate on specific nuances within the training dataset, thereby compromising their generalization capabilities on the validation dataset.

Further scrutinizing these six U-Net architecture and encoder groups, more standout performances were observed from ResNet50, Xception, and EfficientNetB4. Notably, ResNet50 exhibited the most robust learning capabilities on the training dataset, with an impressive average IoU of 86%. Additionally, its performance metrics on the validation dataset were commendable. Meanwhile, EfficientNetB4 demonstrated marginally superior generalization capabilities in the validation dataset. This might be attributed to its specialized architectural tweaks and adjustments, oriented towards effective feature capture and overfitting deterrence.

Thus, we've resolved to further refine the ResNet50 model and EfficientNetB4 model in our subsequent phase. The overarching objective will be to mitigate overfitting to enhance the model's generalization potential.

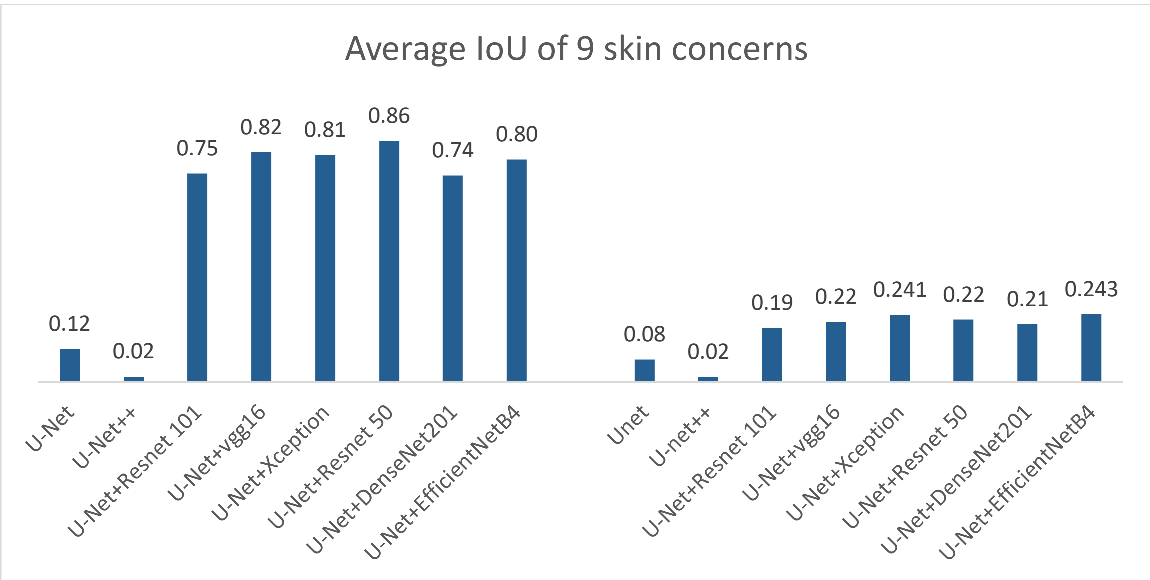
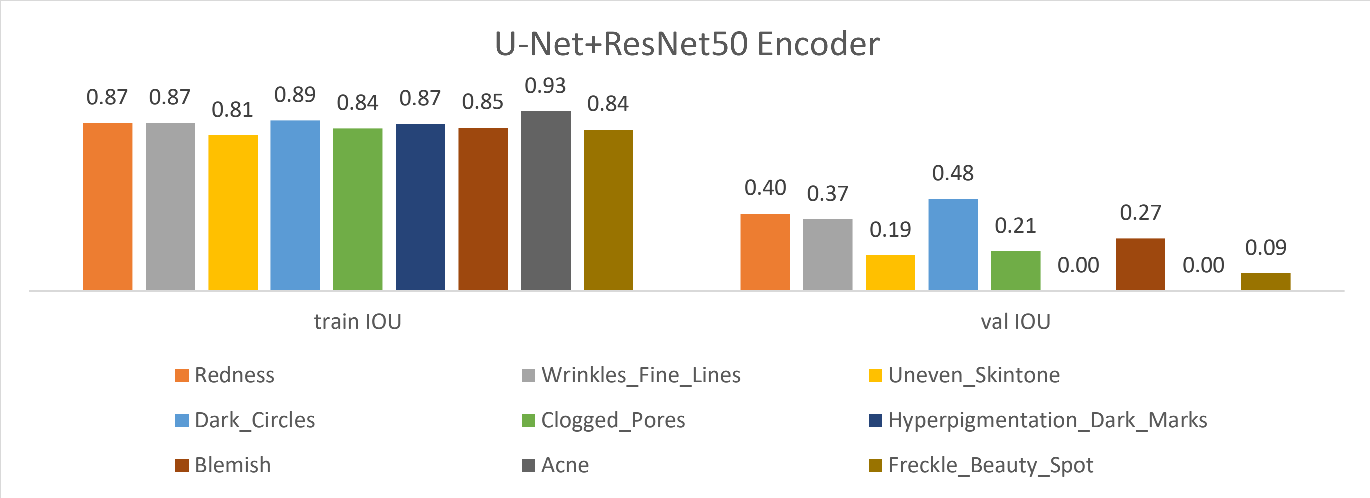


Figure 4: Average IoU of 9 skin concerns across 8 groups (Left side is training set, right side is validation set)
Note: sample size n=327 after data cleaning



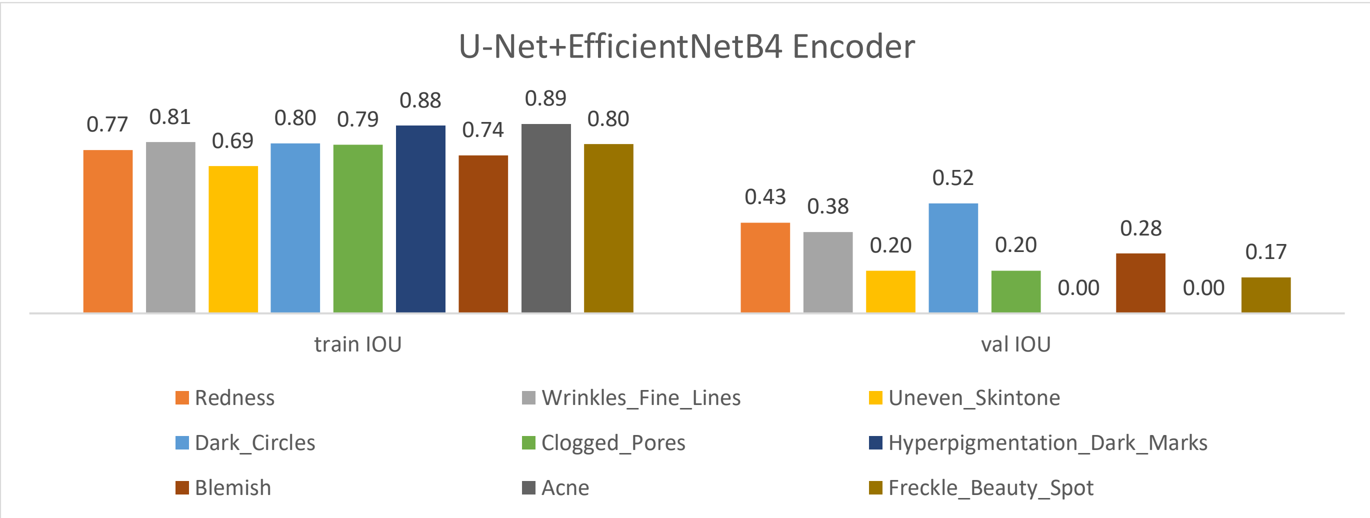


Figure 5: . IoU of 9 skin concerns by difference encoder

Note: sample size n=327 after data cleaning

Looking at the learning and prediction capabilities of the 9 skin concerns specifically in **Figure 5**, ResNet50 and EfficientNetB4 show a similar pattern, i.e., better generalisation for Redness, Wrinkles Fine Lines, Dark Circles, but Hyperpigmentation Dark Marks, Acne, and Freckle Beauty Spot were particularly poor in generalisation.

In summary, ResNet50 and EfficientNetB4 encoder models were selected for next-stage optimization, and the main task is to find the cause and solve the overfitting problem to improve the generalisation of the model, especially for Hyperpigmentation Dark Marks, Acne, and Freckle Beauty Spot skin concerns.

5.2 Overfitting Discussion and Solutions

The previous section presented the initial results of model comparison between ResNet50 and EfficientNet with 327 images. Then, we moved to the second stage of model strengthening by enlarging dataset to address overfitting problem and deploying a series of optimizing strategies such as ensemble learning, regularization, hyperparameter tuning and facial area detections. The following section will discuss the possible reasons behind the overfitting performance and present the eight strategies we utilized to improve model performance.

5.2.1 Increasing Sample Size to Mitigate Overfitting

Both articles by Mutasa et al. (2020) and Ying (2019) delved into the causes and solutions of overfitting. Several solutions to minimize the overfitting effect were proposed, including the most wildly used and effective way, which is expanding the training dataset. In this research, we first adopted the dataset expansion method to mitigate the overfitting effect. The dataset was enlarged from 327 to 1,000 images following the same quality filtering protocols. Both Resnet50 and EfficientNet model were applied to detect the facial images and the performance was measured in mean Intersection over Union (IoU). The model performance comparison based on the enlarged dataset between two models is presented in **Figure *6***.

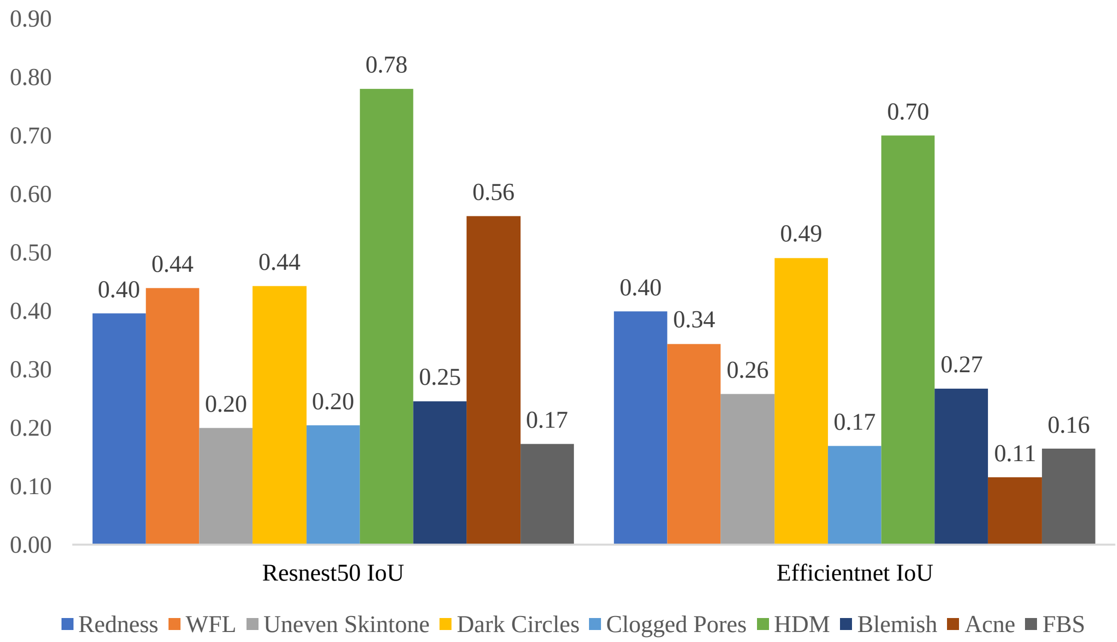


Figure 6:Model performance comparison between Resnet50 with EfficientNet based on test set mean IoU.

Notes: WFL stands for Wrinkle Fine Lines; HDM stands for Hyperpigmentation Dark Marks; FBS stands for Freckle Beauty Spots.

From **Figure *6*** we could observe the ResNet50-based U-Net model exhibits superior performance over EfficientNet across multiple skin issues, particularly in significant skin concerns like Wrinkles Fine Lines, Hyperpigmentation Dark Marks, and Acne. With an average IoU of 0.38, surpassing EfficientNet's 0.32, ResNet50 demonstrated its superior adaptability across a wider range of skin issues. Given its consistent and enhanced accuracy, we decided to adopt ResNet50 as the foundational encoder for further model improvements.

Table 3: Recall results on 1000 samples using ResNet50-based U-Net model.

Sample Number	Skin Concern Class								
	Redness	Wrinkles Fine Lines	Uneven Skintone	Dark Circles	Clogged Pores	HDM	Blemish	Acne	FBS
327	60%	57%	26%	66%	31%	0%	36%	0%	15%
1000	**73%**	**67%**	**32%**	**70%**	28%	0%	27%	0%	**20%**

Notes: HDM stands for Hyperpigmentation Dark Marks; FBS stands for Freckle Beauty Spot.

Table 3 presents the prediction accuracy for each skin concern class after data enrichment. Several skin concerns' accuracy had been witnessed to increase significantly. Among the nine skin concerns, the highest achieved recall is 73% with skin concern Redness, whilst the other calculated values are 70% with Dark Circles, 67% with Winkles Fine Lines, 32% with Uneven Skintone and 28% with Clogged Pores. Compared to the figures of the initial sample size (327 images), we could observe an overall positive impact on the model's learning performance by increasing data volumes.

From our analysis, the model commendably captures specific skin concerns, particularly: Redness, Dark Circles, and Wrinkles Fine Lines. We are optimistic that its performance will further augment with the inclusion of more training data. However, a persistent challenge arises when detecting Hyperpigmentation Dark Marks and Acne, which both consistently recorded a 0% accuracy rate. Delving deeper, we deduced that this shortfall could be attributed to the potential absence or underrepresentation of these skin concern classes in the training dataset.

Table 4: Distribution of the two under-represented skin concerns.

Skin Concern	Among 22,132 Images	1st round of manual screening 14,515 images	1000 sample
Acne	1292	640	61
HDM	1582	1049	110

Notes: HDM stands for Hyperpigmentation Dark Marks.

Table 4 presents the distribution of two imbalanced skin concern classes in Yuty's dataset. According to our research, among 22,132 images of Yuty's database, only 1,292 images contain the Acne problem, and 1,582 images contain Hyperpigmentation Dark Marks problem, which takes only 5.8% and 7.1% of the total dataset. The distribution of the two specific skin concerns was further shrunk in the 1,000 samples, which caused difficulty for the model to learn and correctly identify the features. Therefore, enriching our data with these specific classes and integration of ensemble learning had been studied in our later section.

Moreover, the limited memory and unit constraint of Google Colab had been the technical barrier to our further sample enrichment. Despite upgrading to a pro membership, we continue to grapple with computational constraints. As for our image segmentation analysis, each picture was cut into 16 smaller pieces with same-sized masks, equal to 16,000 small images in the 1,000-sample data. The maximum memory of Google Colab had been fully used.

In our pursuit of data enrichment, it became evident that Google Colab's computational capabilities would not suffice to maximize model performance. As a benchmark, training on every 1,000 images consumed 30 units, which equates to an investment of roughly 3 pounds, and required 60 minutes. Recognizing the need for better outcomes, we took the initiative to self-fund and undertake multiple iterations to refine the results. Our explorations have led us to consider the high-performance computing cluster, Apocrita. While it's still in the process of being integrated, its potential makes it a promising avenue for the future evolution of our model.

5.2.2. Weight Initialization for Imbalanced Skin Concern Classes and Regions

The imbalance between skin concern classes and regions also results in the overfitting problem. Nine skin concern classes along with their counts are shown below. Among the classes, Redness and Uneven Skin Tone have relatively higher counts, with 6,090,981 and 6,675,847 respectively. Conversely, Hyperpigmentation Dark Marks has the lowest count, with a total of 334,355. This discrepancy in the counts between different skin concern classes leads to a problem of imbalance. Furthermore, images with and without skin concerns also show an imbalance problem. The average skin concern area is 13,848 (2.3%) while the average non-skin concern area is 575,975 (97.7%).

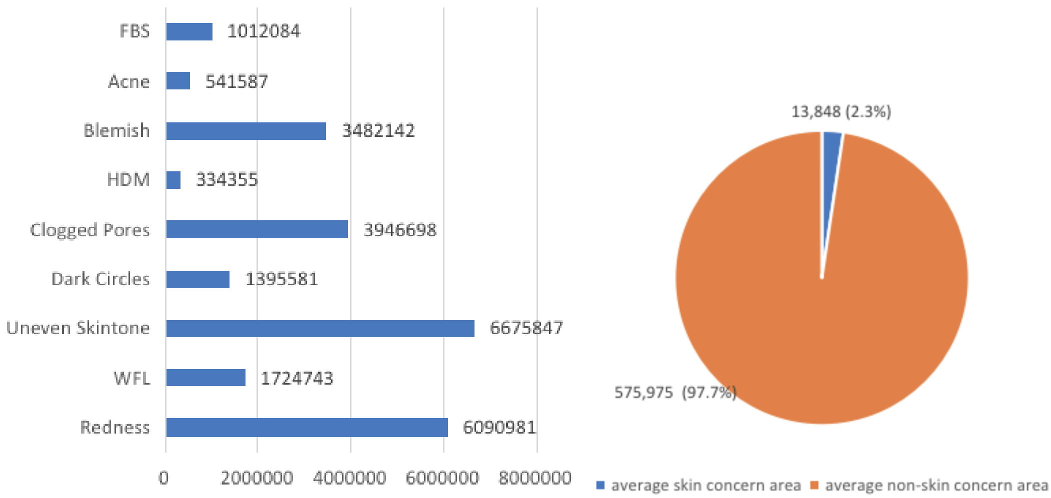


Figure 7: Comparison of skin concern classes and regions.

Notes: WFL stands for Wrinkle Fine Lines; HDM stands for Hyperpigmentation Dark Marks; FBS stands for Freckle Beauty Spots.

To address this issue and alleviate the overfitting problem, we need to initialize different weights to nine skin concern classes and regions to balance the sample size.

codes:
class_weights = 1. / np.array(problem_counts)
skin_problem_weight = average_non_skin_problem_area / average_skin_problem_area

Our model defines focal loss by assigning more weight to the underrepresented skin classes. Our model gives higher weights to classes that appear less frequently and lower weights to those that appear more frequently. The second line of code determines the weighting factor for skin concerns based on their mean spatial coverage within the images, which is particularly useful when the area of interest (the skin concern) is much smaller than the background or other non-concern areas. Before running the above codes, the number of skin concern pixels and the total skin problem area had been computed in advance.

Table 5: Model performance before and after weight initialization.

Method: Weight Initialization	Mean IoU	Skin Concern Class (Count)									Average
		Redness 6090981	WFL 1724743	Uneven Skintone 6675847	Dark Circles 1395581	Clogged Pores 3946698	HDM 334355	Blemish 3482142	Acne 541587	FBS 1012084	
Original	train	0.65	0.72	0.72	0.73	0.64	0.34	0.72	0.66	0.48	0.63
	test	0.22	0.33	0.11	0.35	0.11	0.00	0.20	0.00	0.03	0.15
After weight initialization	train	0.85	0.89	0.82	0.85	0.79	0.9	0.83	0.89	0.89	0.86
	test	0.37	0.4	0.16	0.48	0.17	0.00	0.24	0.00	0.13	0.22

Notes: WFL stands for Wrinkle Fine Lines; HDM stands for Hyperpigmentation Dark Marks; FBS stands for Freckle Beauty Spots.

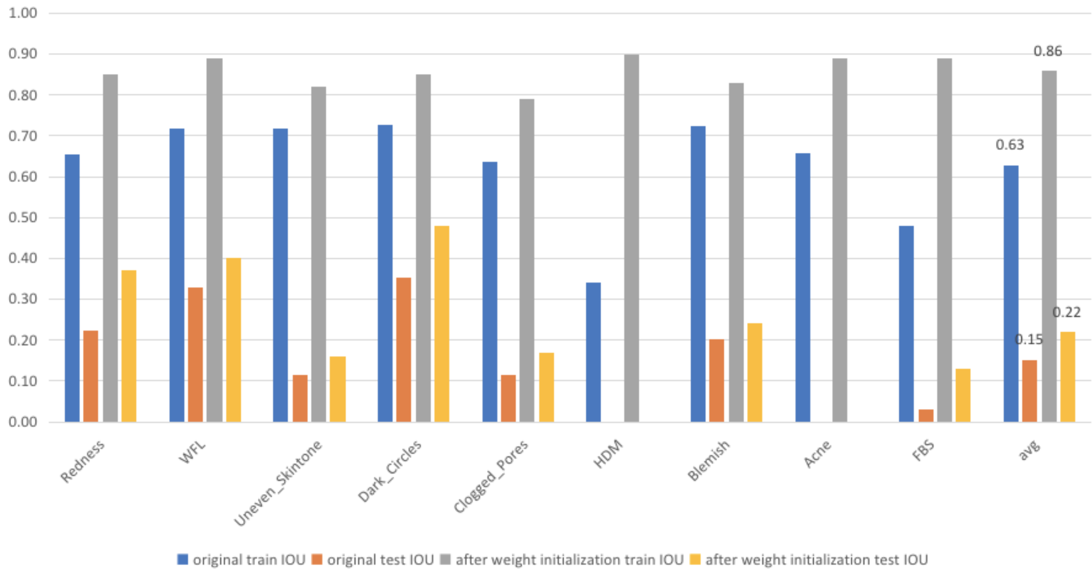


Figure 8: Model performance before and after weight initialization.

Notes: WFL stands for Wrinkle Fine Lines; HDM stands for Hyperpigmentation Dark Marks; FBS stands for Freckle Beauty Spots.

According to the table and graph above, after applying the weight initialization method, the average mean train IoU increased from 0.63 to 0.86 and test IoU increased from 0.15 to 0.22. The mean IoU values generally increased for most skin concern classes. This indicates that the model's segmentation accuracy improved after addressing the imbalance issue through weight initialization. The improvement in mean IoU is more noticeable in some classes (e.g., Redness, WFL, Dark Circles, Clogged Pores, HDM, Blemish, Acne, and FBS) both in the training and test sets.

5.2.3. Data Augmentation

One of the most classical and modern remedies for overfitting problems is data augmentation (Rice, et al., 2020). Firstly, based on the possible Yuty user scenarios in reality, we define four data augmentation methods: random rotation (-10, 10) degrees, random translation (-50, 50) pixels, random horizontal flip, randomly adding gaussian noise and gaussian blur.

Next, we divide the filtered images into four copies in order. One of four data augmentation methods is applied for each copy of the images. Afterwards, we update the coordinate information in the JSON file, output the shape information of each image after augmentation, and save the augmented images and the updated JSON file. Finally, we combine the augmented images and filtered images as a new folder, merge the original JSON file and augmented JSON file into a new one, and then perform model training as before. Take 500 images as example, the manually screened and quality filtered images for 500 images is 175 images, which will generate 172 augmented images and a combination of 347 images in total for model training.

Table 6: Model performance on data augmentation (ResNet101 as the encoder).

Method: Data augmentation	Mean IoU	Average
Original	train	0.732
	test	0.170
Augmented data	train	0.710
	test	0.080

From the table above, we can see that the mean IoU values in both the training and test sets do not improve, but rather decrease after data augmentation, which is not what we expected. The reason why data augmentation cannot improve our model performance might be the difference between the given data quality and the data quality in real Yuty user scenarios. Specifically, the given images in the dataset are rarely rotated, transformed, or flipped, but Yuty users may upload rotated, transformed or flipped images in real use scenarios. If we use images in real scenarios for training and test, our model performance is expected to improve, and the overfitting problem should be alleviated to some extent after performing data augmentation.

5.2.4 Hyperparameter Tuning to Mitigate Overfitting

Hyperparameter optimization remains pivotal in enhancing the efficacy of machine learning models. We can establish a delicate equilibrium between training precision and model robustness by fine-tuning these parameters. Such a balance is instrumental in reducing the detrimental effects of overfitting—characterized by the model's undue reliance on training data—and in averting the shortcomings of underfitting, where the model becomes too generalized, sacrificing detail. To bolster our model's performance, we have employed the subsequent strategies:

- *Regularization: L2 Regularization*

Researchers have devised numerous innovative strategies to optimize the use of limited training data and curtail overfitting. Mutasa et al. (2020) suggested modifying the learning algorithm to encourage the model performance, also known as regularization.

One of the regularization techniques we used in this research is L2 Regularization (Ridge Regression). As in the learning process of neural networks, each feature has been given a weight to signify their importance to ensure that no particular feature excessively dominates others, and only features that significantly improve the initial cost function are assigned with large weights.

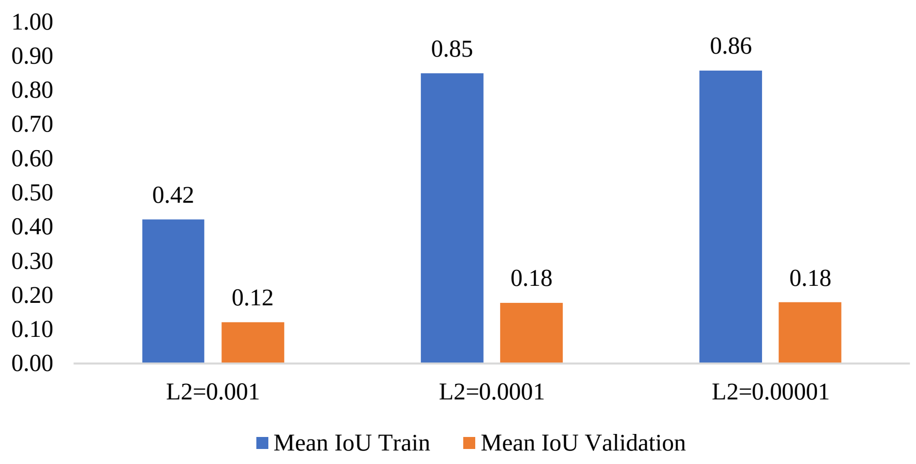


Figure 9: Comparing model performance with modified L2 regularizations.

Notes: The results of model performance using modified L2 regularization from 0.001 to 0.00001 are presented in Table 3. We used the 327-sample package for better quality and lower data noise as this package was selected by two rounds of manual screening and quality filtering. Both the training and validation dataset had been evaluated by the mean Intersection over Union (IoU).

We could observe from **Figure *9*** that when L2 equals 0.0001, the overall performance for the validation set achieved the best result, combined with a slight improvement in mitigating the overfitting effect in the training dataset. Reducing L2 from 0.0001 to 0.00001 did not show a significant improvement in model performance and might weaken the regularization effect. Given this observation, we decided to retain the L2 regularization value of 0.0001 to ensure greater stability and robustness in our model.

- *Regularization: Dropout*

Another technique we used for regularization is Dropout. This approach is a prevalent countermeasure to overfitting in modern neural network models. Its core principle involves the random omission of units and their associated connections during training, which prevents units from becoming excessively interdependent. According to Ying (2019), Dropout not only suppresses overfitting but also enables the combination of predictions from diverse, extensive neural networks during testing, also considerably trims computations amounts.

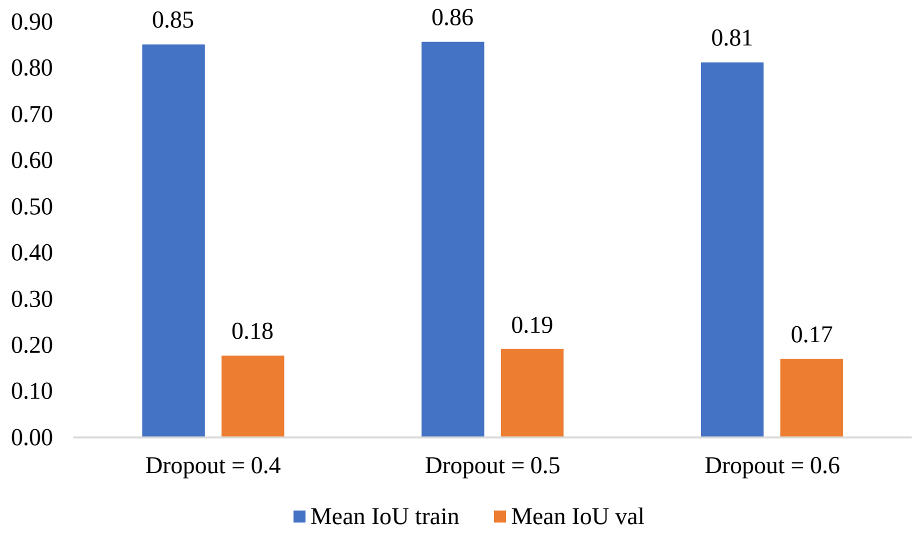


Figure 3. Comparing model performance with modified Dropout rate.

From Figure 3, we observed a dropout rate of 0.6 effectively mitigated overfitting, while also led to a marginal decline in the validation mean IoU, from 0.18 to 0.17. Compared to the model performance in Figure 2, there was no noticeable impact of adjusting Dropout rate with validation dataset's results maintained the same or improved. Ying (2019) also mentioned the possibility that Dropout might reduce the model's capacity to capture features and lead to performance degradation and inconsistency in unseen data. Therefore, we left the dropout rate and continued the model with only L2 regularization.

- *Neural Network Architecture (decoder simplification)*

From previous experience, we found the growth in the network's size could lead to a surge in computational resource usage and present challenges in optimization. In this light, Zhang et al. (2023) discussed the impact of dense connections within the U-Net-based single and combined models. The research discovered that excessive dense connections could lead to overfitting, even though there may be an enhancement in Accuracy.

Therefore, we systematically experimented with reducing the number of decoders in our model to compare the model performance and investigate whether such an approach could alleviate the overfitting effect.

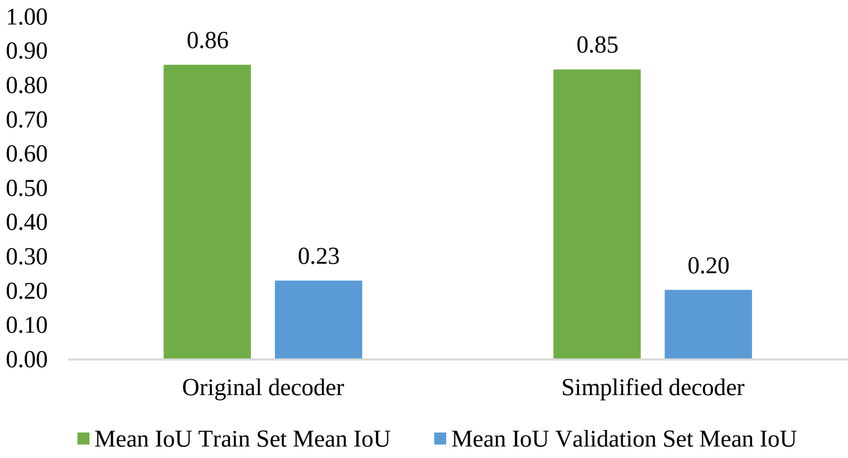


Figure 10: Model Performance Comparison with Simplified Decoders.

The results of model performance based on simplified decoder arepresented in *Figure 10*. The new decoder begins with fewer filters (256 instead of 512) and has fewer layers overall, which shows a simplified structure compared to the original model and enabled the modified decoder faster in terms of computation and demand less memory.

However, a simplified decoder also suggests a reduced capacity and may not be able to capture as much detail or complexity from its input as the original one. As we could observe from the graph, the simpler structure reduced the validation set's performance from 0.23 to 0.20 where the overfitting effect was still strong. Therefore, we decided to maintain the original decoder structure as Yuty has nine skin concerns and would require a more complex model to identify the features. Meanwhile, the zero mean IoU on the two specific skin concerns (Acne and Hyperpigmentation Dark Marks) were due to the heavily under-represented class dataset that we aimed to improve in later steps.

5.2.5 Other Means to Mitigate Overfitting

- *Using a combination metrics of Accuracy, Dice Coefficient, and Recall.*

After enlarging the training dataset and improving model performance with hyperparameter tuning, we have introduced several other ways to mitigate overfitting problem. For example, we added two more metrics - Dice Coefficient and Recall- besides the original Accuracy to better evaluate the model performance. Although Accuracy is a global metric, it could cause misleading results when classes are imbalanced. Some skin concerns with more extensive data might overshadow the rarer ones regarding Accuracy, making the metric less informative about how well the model recognizes the underrepresented skin concerns. Adding Dice Coefficient could emphasize the quality of the segmentation overlap for each class, and Recall ensures that even the rarer skin concerns are being detected. Also, Recall would be used as our primary metric as a measure of a model's ability to correctly identify all relevant instances, which is especially important in medical or skin concern contexts.

- *Adjust thresholds for under-represented classes.*

Moreover, as the model might be biased towards the majority class, adjusting the thresholds can help achieve a better balance between classes. Adjusting the thresholds from default number 0.5 to a lower degree (e.g., 0.3) would prioritize not missing any potential skin concerns, which might be preferable to Yuty to flag potential skin issues for further examination rather than miss them. On the other hand, increasing the thresholds to above 0.5 would make the model stricter in its positive predictions. In our case, we further investigated the impact of altering thresholds on the imbalanced dataset's performance improvement.

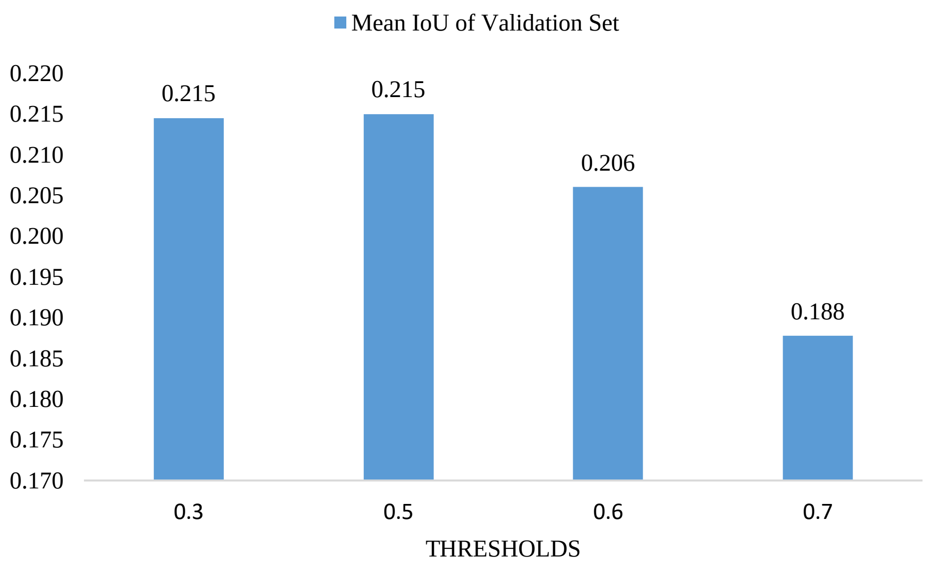


Figure 11: Comparing model performance with modified thresholds on validation test set.

The results of model performance on the modified threshold are presented in **Figure *11***. We could observe a slight increase in the validation set when the threshold was altered from 0.5 (default value) to 0.3. Such an approach could make the model more lenient in predicting positive cases and create a higher propensity to capture more true positive cases (higher Recall).

However, we did not observe a significant improvement in the imbalanced classes' performance (Acne and Hyperpigmentation Dark Marks), which may require further research after enriching the training datasets of these two specific skin concerns.

5.3 Model Enhancement: Ensemble Learning & Model Stacking

After data enrichment and hyperparameter tuning, we then moved to the next step which is tackling the imbalanced class problem.

According to Li et al. (2019), when trained on limited data, the model struggles to correctly classify new samples from classes that were under-represented in the training data as the less-represented class tends to gravitate towards, or even cross, the decision boundary, adding more uncertainties in the decision-making. Meanwhile, the dominant class remains relatively stable. This finding has clearly explained the low recall figure of Acne and Hyperpigmentation Dark Marks, as both skin concerns are heavily under-represented in Yuty's dataset. To address this, we implemented 'ensemble learning' to optimize performance on imbalanced skin concerns. Essentially, an ensemble of models trained on different subsets of data to reduce the variance in predictions. Studies of Snider et al. (2023) have also revealed the tactic of using data augmentation and ensemble prediction to prevent machine learning overfitting.

Subsequently, we developed a supplementary model focusing on the Acne feature. After training it with 579 Acne-enriched images, both models were tested on the final dataset. The secondary model is anticipated to offer a more dependable prediction for this under-represented class.

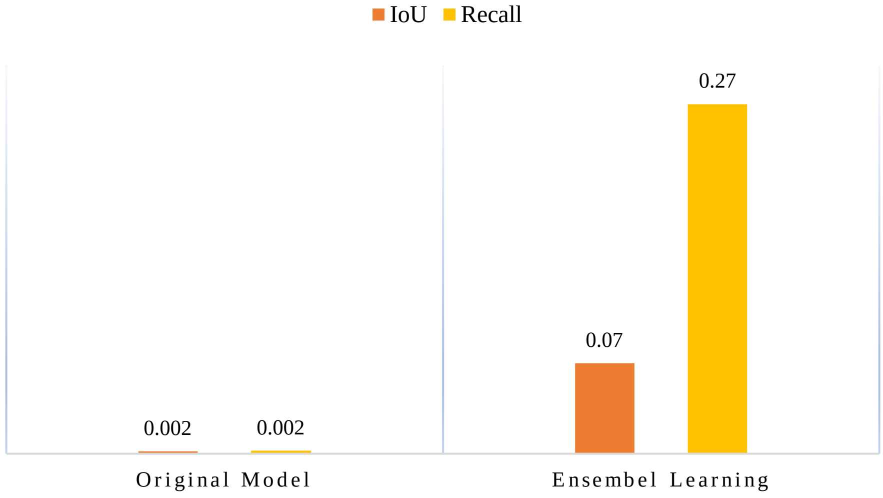


Figure 12: Enhanced Acne results with ensemble learning.

Notes: The second model is an Acne enhancement model that specially trained with 579 Acne sample images that are collected from Yuty's database and processed with two rounds of manual screening and quality filtering. Model 2 is then used to predict with the final test dataset to verify the effectiveness of ensemble learning.

Table 7: Model performance based on ensemble learning.

Validation set		Redness	WFL	Uneven Skintone	Dark Circles	Clogged Pores	HDM	Blemish	**Acne**	FBS	Average
		Skin Concern Class									
Ensemble Learning	Mean IoU	0.45	0.42	0.16	0.47	0.23	0.03	0.27	**0.07**	0.14	**0.25**
	Recall	0.65	0.60	0.19	0.58	0.37	0.03	0.39	**0.27**	0.20	**0.36**

Notes: WFL stands for Wrinkle Fine Lines; HDM stands for Hyperpigmentation Dark Marks; FBS stands for Freckle Beauty Spots.

From Figure 12 we could observe a significant increase in Acne's mean Intersection over Union (IoU) from 0.002 to 0.07 signals an improvement in our model's accuracy in making predictions that closely match the actual data. Meanwhile, the jump in Recall from 0.002 to 0.27 shows the model is now better at correctly identifying true instances without missing many.

Overall, the ensembled model that specialized in the imbalanced dataset had successfully leverage the model performance in detecting imbalanced skin concerns, and we expect such improvement continue to increase when more training data are included. Table *7*, showed the detailed performance for each skin concern. The second Acne enhancement model has leveraged the average IoU from 0.22 to 0.25 compared to the initial results in detecting across the nine skin concerns.

During the experiment, we also noticed the computational constraints from Google Colab. The maximum samples we could use is 1000 images from the original nine skin concerns model and 579 Acne sample for ensemble learning. Then we would hit the maximum system RAM that Google Colab could provide. In practical scenarios, certain skin conditions are encountered less frequently. Given the demonstrated performance uplift achieved through ensemble learning, we advocate for the integration of specialized architectures tailored to these infrequent skin concerns. Additionally, prioritizing a larger data collection for underrepresented classes can further enhance model performance.

5.4. Facial Region Detection to Correct Wrong Predictions Outside the Face

So far, the models have been making predictions on the entire selfie, not just on the face, and this can pose a serious problem, especially when our models are aimed at the end consumer, when the predicted skin problems appear in background areas etc., take *Figure 13* left picture for example, and the accuracy of the model will be seriously doubted.

Therefore, we need to first stitch the 16 small photos that were cut into 256*256 back to the original large picture of 1024*1024. Use the pre-trained face detection model to extract the face area of the selfie, and finally cross-check the skin problems predicted by the model and the face area, and delete the skin problems predicted outside the predicted face area. *Figure 13* is an illustration of the process and effect. By applying face detection model, the average IoU of the prediction increased 0.07% as shown in *Table 8*.

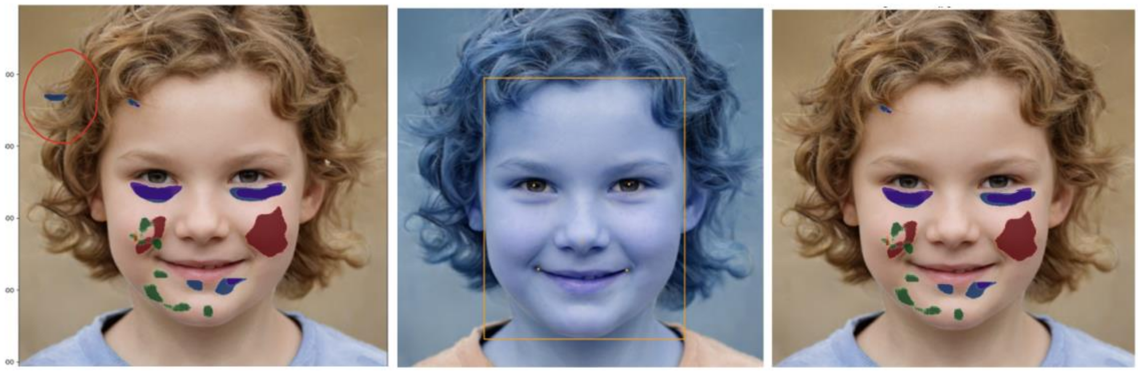


Figure 13: Comparison after apply Face recognition model to correct the prediction result (The left picture is the original prediction result, and the right side is the result after combining MTCNN facial recognition, which shows that the skin problem prediction on the hair on the left is deleted)

Table 8: Prediction uplift after apply face detection model

	Avg IoU in Validation dataset
without face detection	18.35%
corrected after face detection	18.42%
Uplift	+0.07%

5.5 The Optimal Model Performance

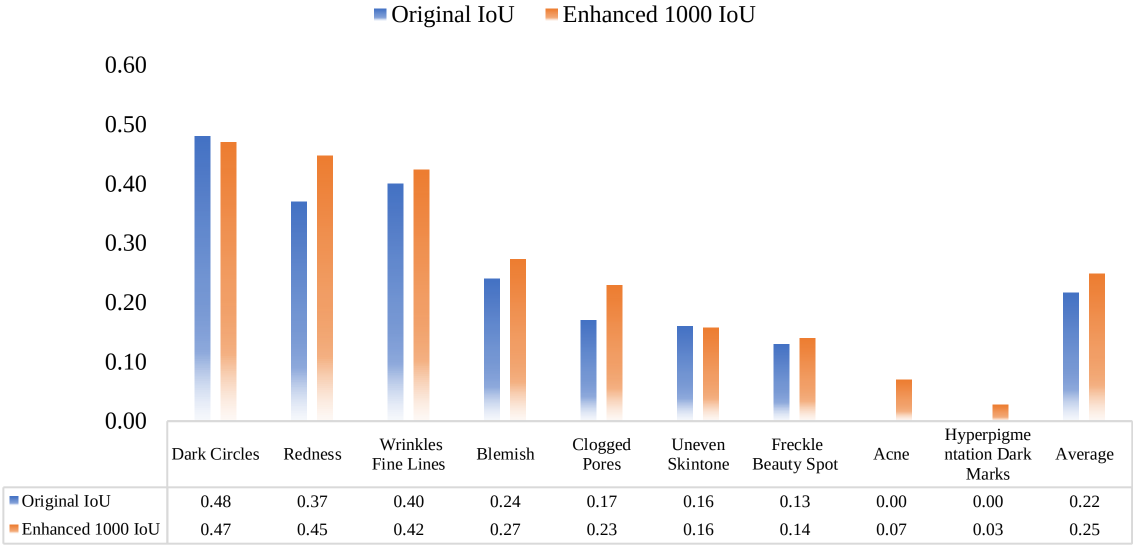


Figure 14: Enhanced overall model performance by intergrating experimented optimising stretegies.

Figure 7 showcases the significant advancements made from our original to our enhanced model. Originating from a limited dataset of 200 samples, the initial model encountered challenges with imbalances, most prominently in its inability to detect Acne and Hyperpigmentation Dark Marks, both of which had a 0.00 IoU.

To tackle these limitations, our enhanced model incorporated a broader 1,000-sample dataset and introduced a specialized Acne model trained on an additional 579 samples. Beyond this expanded data approach, the progress also stemmed from a series of meticulous experiments targeting the overfitting issue. These included the introduction of L2 Regularization, Dropout techniques, and a streamlined decoder design. Further, our evaluative metrics evolved; we integrated a composite measure combining Accuracy, Dice Coefficient, and Recall to yield a nuanced performance analysis, facilitating more precise threshold adjustments and an addition of facial area detections.

This robust methodological refinement was evident in the outcomes: the average IoU elevated from 0.22 to 0.25. Summarizing, the chart not only reflects our model's tangible progress but also underscores the strategic interventions and diligent experimentation that drove these enhancements.

Table 9: Tuned Hyperparameter for optimal model performance.

Optimizer	gamma	alpha	learning_rate	L2	batch size	validation split	epochs
Adam	0.2	0.2	0.0001	0.0001	32	0.1	100

Table 10: The optimal model performance after a series optimizing strategies.

	Skin Concern Class									
	Redness	WFL	Dark Circles	Blemish	Clogged Pores	Acne	FBS	Uneven Skintone	HDM	Average
IoU	0.45	0.42	0.47	0.27	0.23	0.07	0.14	0.16	0.03	0.25
Dice Coefficient	0.62	0.59	0.64	0.43	0.37	0.13	0.25	0.27	0.05	0.37
Recall	0.65	0.60	0.58	0.39	0.37	0.27	0.20	0.19	0.03	0.36

Notes: WFL stands for Wrinkle Fine Lines; FBS stands for Freckle Beauty Spots; HDM stands for Hyperpigmentation Dark Marks.

Table 10 evaluates our final model's performance across all nine skin concerns, using three key metrics: Intersection over Union (IoU), Dice Coefficient, and Recall.

At the forefront, the model manifests a robust aptitude in detecting Wrinkles Fine Lines, Redness, and Dark Circles. Their leading scores across all metrics underscore this capability. From a commercial standpoint, accurately identifying these prevalent skin concerns aligns with consumers' primary grievances. Being adept at accurately pinpointing these issues translates to the potential of delivering precise skincare recommendations, thereby augmenting customer satisfaction.

However, while categories like Blemish, Clogged Pores, and Uneven Skintone show respectable results, there is evident room for enhancement in Acne, Freckle Beauty Spot, and notably HDM (Hyperpigmentation Dark Marks), which had an IoU of 0.00 due to the lack of specific class instances in training samples. This could be improved by future dataset enrichment. The high Dice Coefficient for HDM, in contrast to its IoU, suggests that while the precise overlap might be a challenge, the general area prediction is commendable.

Significantly, our refined model boasts a 14% increase in IoU compared to the initial results. This tangible progress speaks to the efficacy of our optimization techniques. We are optimistic about further enhancements with more extensive training data.

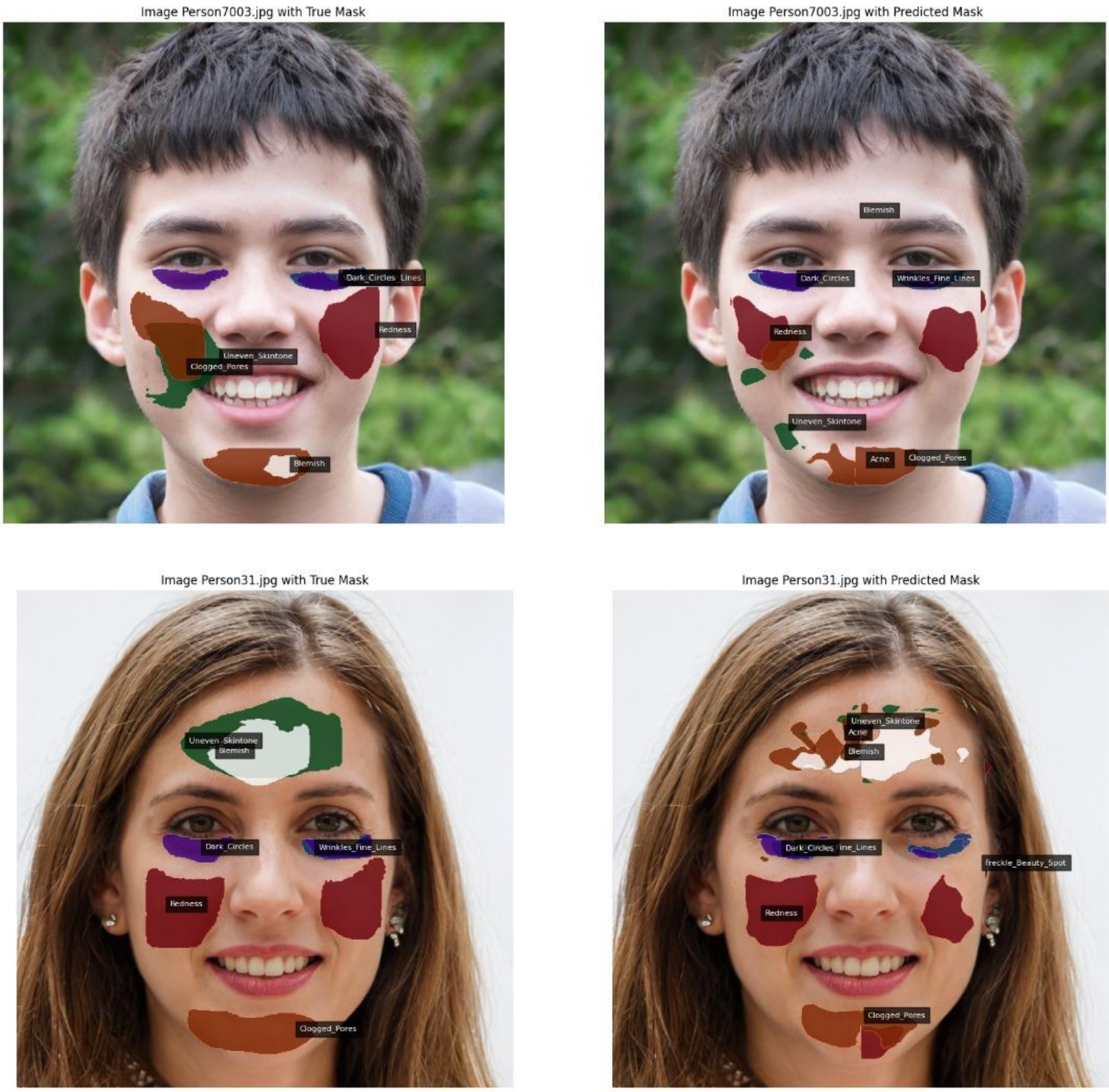


Figure 15: Examples of actual results and our predicted results.

Figure 15 presents a side-by-side comparison of the original image masks with the model's predictions. A close alignment between the two is observed, indicative of a commendable performance in skin concern detection.

In summation, the model's prowess in discerning primary skin concerns augurs well for business expansion and customer rapport. Nonetheless, a continued commitment to refining the model, especially in its weaker areas, remains essential to ensure holistic and dependable skin analyses for a diverse user base.

5.6 Limitations & Further Improvements

In our study, we proposed and optimized a ResNet 50-based U-Net model that was specially designed for Yuty's nine skin concerns facial image segmentation project. Our architecture is proven efficient in capturing three specific skin concerns: Redness, Wrinkles Fine Lines and Dark Circles with limited training datasets. However, we also encountered a few limitations during model development. We have summarized the detailed limitations as below and provided alternative solutions for future improvements:

I. Limited training data input due to the constraint of Google Colab memory and unit upper limit, which could be improved by using more powerful computing tools such as a high-performance cluster like Summit or Sierra in the future.
II. Heavily under-represented data in specific skin concerns (Acne and Hyperpigmentation Dark Marks) led to a hamper in learning capability and prediction accuracy in unseen data. The alternative solutions could be collecting more relevant data samples, producing new data based on the current database, or using a pre-trained model on the specific skin concern to improve the model's performance in the future.
III. The third challenge was splitting the train, test, and final test dataset with a proper distribution of each skin concern assigned accordingly. Such a task is especially challenging in image segmentation as one image may have several skin concerns, and some features are naturally rarer to see when data are randomly selected. Splitting such data while preserving an equal distribution in all subsets is inherently difficult.
IV. In our quest to develop a deep learning model capable of identifying a wide range of demographics and skin types, we acknowledge the challenge posed by our limited sample size. This constraint may affect the model's ability to fully accommodate variations in lighting conditions, skin tones, and facial expressions. Consequently, efforts to enhance the diversity of our dataset are paramount.
V. Another limitation is the required technology to zoom in and detect small/unobvious skin problems. As this knowledge is quite cutting-edge and still progressing, we need more studies and investigations in the related areas to enhance the model performance further.
VI. The last limitation we encountered was the data noise. For facial skin image segmentation, the amount of skin concern data at the pixel level is typically small,

making the annotation data's accuracy even more vital in the model's correct prediction. In Yuty's case, the skin concern annotations were marked by AI, and many images were noisy-labelled, which caused the disturbance and ineffective training results in skin concern identification. We manually rechecked two hundred images with the marked skin problems and found a few with unmatched results. As suggested by Abbas et al. (2011), Yuty could also consider using a more judiciously annotated dataset marked by dermatologists in the future for better model performance.

Notably, while the U-Net architecture has been widely recognized for its proficiency in medical image segmentation, it can occasionally exhibit susceptibility to overfitting, particularly in scenarios where the available dataset is limited. Integrating an extensive encoder like ResNet50, although beneficial for capturing intricate patterns, might inadvertently accentuate this tendency towards overfitting. In light of these challenges, it might be constructive to explore alternative architectures, such as the Multi-scale Dense U-Net or the Attention U-Net. Such architectures have shown promise in managing overfitting by leveraging multi-scale features or by focusing on salient regions of the input, respectively. Their inclusion could offer a balanced trade-off between complexity and generalization, potentially leading to enhanced model performance.

In addition, exploring noise reduction techniques could further refine our model. Drawing inspiration from Zhu et al.'s 2019 research, the prospect of deploying a network that critically evaluates label quality presents a compelling direction to explore. Moreover, addressing overfitting through methods like pruning—removing extraneous or less relevant data—could further bolster both the efficiency and accuracy of our model. These approaches present intriguing avenues for our forthcoming research.

6. Business Implications

As highlighted in the introduction, the burgeoning demands of the industry have rendered the product matrix more intricate and diverse, leading to a paradox of choices for consumers. Research from Wakefield Research (2018) illustrated that '70% of beauty consumers feel overwhelmed due to the vast array of beauty product choices and 63% of consumers find beauty product claims confusing.' Consequently, consumers grapple with selecting the right product and making informed purchase decisions. This congested environment has escalated the demand for personalization. Moreover, data indicates a significant appetite for personalization within the beauty and skincare industry: 71% of shoppers seek a personalized shopping experience, and over 80% are more likely to revisit a brand offering personalized services (McKinsey and Company, 2021).

Aligning with this market trajectory, Yuty anchors its objectives in these consumer demands. As an emerging enterprise, Yuty is on the brink of amassing $1 million in revenue by 2023. Its unique selling proposition is the deployment of sustainable artificial intelligence to curate personalized product recommendations, fostering a seamless shopping journey (Kanetkar, 2021).

Pertaining to our project, we integrated AI-driven image semantic segmentation technology. After examining a myriad of skin concerns, we successfully realized the detection of nine facial skin issues. Moving forward, Yuty can synergize the model's detection outcomes with its established personalized recommendation system. This creates a comprehensive commercial chain, spanning from consumer selfie uploads, skin issue diagnosis, to dispensing tailored advice and product suggestions. A study by Forrester Research posits that entities utilizing AI to proffer personalized client experiences can foresee a 10% uplift in sales. As a trailblazing product in the market, this project will further refine Yuty's AI-driven personalized product portfolio, catalyzing its business growth.

From a SWOT vantage point, we analysed Yuty's position and deliberated on how our project could amplify Yuty's strengths, mitigate its weaknesses, and navigate potential challenges to realize its financial aspirations.

However, when deploying this model, it's crucial to acknowledge its limitations. Due to data constraints, prediction accuracy varies slightly across the nine skin issues. Given that accuracy significantly influences consumer confidence in our product and, subsequently, brand perception and loyalty, we recommend strengthening the model's accuracy by augmenting the data volume. Until a larger dataset refines the model, an initial focus could be on issues with the highest prediction accuracy, such as Redness, Wrinkles, Fine Lines, and Dark Circles.

6.1. Strengths

6.1.1. Current

Yuty has a unique brand positioning which is built on strong values of trust and transparency. Even though it’s a new company, Yuty was recognised as one of the "Most Innovative Beauty Brands" by The New York Times in 2021. The company has been praised for its commitment to sustainability and its use of AI to personalize beauty routines which have helped the company to attract a loyal customer base. Yuty's machine learning model is a unique product offering where the company's AI can analyze a user's lifestyle, skin concerns, climate, etc. to recommend personalized products and skincare routines through a questionnaire. This gives the company a competitive edge over its industrial rivals.

6.1.2. With respect to our project

In this highly competitive market, no current market offerings are quite like our product. By incorporating this product, we can further refine the product matrix, making its personalized services more comprehensive and holistic. This is mainly manifested in three aspects--enhancing customer satisfaction, increasing repurchase rates, and minimizing customer maintenance costs. By providing customers with accurate and personalized information about their skin concerns, Yuty can help them to feel more confident in their skin and to find the right products to address their needs. This can lead to increased customer satisfaction and loyalty which in turn can lead to referrals from satisfied customers. Increased satisfaction for skincare products can lead to increased repurchasing, as people often need to use them on a regular basis. With the model's accurate diagnosis and customized recommendation, the customer support tickets will be reduced, which will ultimately lower the costs for the company. Since a study by the National Retail Federation suggests that 72% of consumers are more likely to purchase products from a company that offers personalized recommendations, incorporating this model can also enhance brand reputation.

6.2. Weaknesses

6.2.1. Current

As Yuty is an emerging company, it has low brand awareness in the market which means that only a small portion of potential customers are aware of Yuty's products or services. This limits the company's ability to reach a broader audience and tap into new market segments. Due to its smaller customer base, the company can have a hard time gaining market share and customer loyalty as compared to its well-established competitors.

6.2.2. With respect to our project

Yuty is capable of leveraging technology to heighten its reputation, which paves the way for new market opportunities, lures more clients, and establishes ties with a broader customer base. First, by offering a self-service tool that analyses skin concerns, Yuty can engage customers interactively and increase brand trust. This can foster a positive relationship between Yuty and its customers and drive word-of-mouth referrals. Second,

Yuty can target its marketing campaigns more effectively by tailoring messages to specific skin concerns, increasing the chances of conversion, customer retention and expansion of its consumer base. This can especially benefit the company since currently no established skin concern AI detection tool is in the market and our model can cater to the popular trend of personalization. Given the current audience for personalized beauty products, the company should leverage social media to target young adults and adults, as they have higher acceptability of new things. As this is a new AI product, Yuty can focus on its fun and interactivity, so as to spread more on the Internet, get more social buzz, and win more attention, especially for emerging companies like us, which allows for a more efficient use of marketing budget.

6.3. Opportunities

6.3.1. Current

As a result of the augmented demands, the personalized beauty market is expected to grow by 20% by 2025, reaching a value of $100 billion. As in 2022, the skincare segment accounts for more than 38% of market share of the personalized market, making it the largest segment (Grand View Research, 2019). Brand collaborations are trending and have turned out to be beneficial for all the parties involved (Kalafatis et al., 2012; Kim et al., 2014) and whirl up a talk amongst its target audience, hence making it another marketing tool.

6.3.2. With respect to our project

The demand for personalized skincare is multiplying which proves as a perfect opportunity to launch an operational model which can detect and highlight skin concerns on selfie images. The company can get a first mover advantage by incorporating our model and launching a relatively mature AI technology for its users while promoting it on various marketing platforms, especially social media, to become 'talk-of-the-town' and enlarge its consumer base. The new model can enhance the company's current model and can act advantageous for other brands to partner with Yuty. The partner brands can attract new customers with Yuty's advanced technology application and increased sales, whereas for Yuty, it can partner with big brands which have an audience that it wants to target, resulting a boost in sales for Yuty, enhanced brand awareness and gaining a wider customer base.

6.4. Threats

6.4.1. Current

Yuty has stepped into a market where its competitors are enormous and established brands which pose a threat to the company as they have more resources, experience, brand reputation and huge loyal customer base. Moreover, other brand that are venturing into similar initiatives also pose a threat due to their speed and technique of developing new tools and products. Additionally, due to Yuty's technology-based operations, its customers can raise privacy concerns which can be challenging to tackle for the company. Alongside, government can add to or change the existing regulations which can restrict the company's operations. The machine learning model that Yuty uses is complex and could be susceptible to security breaches. This could damage the company's reputation and its ability to compete.

6.4.2. With respect to our project

In a competitive and technology-driven market, it is important to stay cautious of threats. Incorporating an AI-powered skin concern detection system provides Yuty with a competitive advantage in the beauty industry. It positions Yuty as an innovative and cutting-edge brand, attracting new customers and setting itself apart from competitors. The company should leverage our model to gain insights into broader skincare trends and identify emerging skin concerns that may not have been previously recognized. This can guide Yuty's long-term product development strategy, tap into unexplored opportunities and strengthen its position as a market leader. As for the data security and privacy concern, Yuty should address privacy concerns by being transparent about how it uses its customers' data and placing security measures to protect customer data. It should also try to anonymize the data being collected through the AI model to cater to privacy concerns. It should be prepared for government regulations and work with regulators to ensure that its business is compliant with the law. In addition, Yuty should safeguard its intellectual property, such as its machine learning model, to prevent rivals from duplicating its technology.

As the personalized beauty sector flourishes, Yuty stands poised to excel. Leveraging advanced technologies and regularly reassessing its SWOT ensures Yuty remains responsive to consumer needs, solidifying its leadership in the market.

7. Conclusion

As the beauty industry continues its ascent, the complex matrix of choices available to consumers increasingly underscores the imperative for personalization. Yuty's initiative, anchored in cutting-edge AI technology, marks a significant stride in addressing this demand. By harnessing the power of the ResNet50 U-Net model, we've pioneered a system that can discern multiple facial skin concerns with notable accuracy. This venture promises not only to alleviate the "paradox of choice" that consumers grapple with but also to amplify the user experience, guiding them towards bespoke skincare solutions tailored to their unique needs.

From eight experiment groups, each armed with varying architectures, our research unveiled the potential of the ResNet50 based U-Net model, which delivered an accurate detection in prominent skin concerns, especially Wrinkles Fine Lines, Redness, and Dark Circles, and has demonstrated a 14% improvement in IoU compared to its initial version after we tested a gamut of eight distinct enhancement techniques to fine-tune its prowess.

Despite these advancements, challenges in data representation, data noise and the constraints of technological platforms offer avenues for further refinement. Addressing these challenges presents an opportunity not only for model improvement but also for Yuty to solidify its standing in the beauty industry, driven by data and technology. Meanwhile, this study could serve as a valuable reference for the beauty sector in developing remote dermatology consultation and recommendation system.

The commercial implications of our AI-centric approach are profound. By facilitating a nuanced diagnosis of various skin concerns, Yuty can not only provide consumers with a personalized shopping experience but also potentially boost sales, echoing Forrester's projection of up to a 10% uplift. However, while the horizon gleams with opportunity, the road is fraught with challenges, including formidable competition and the ever-looming concerns surrounding data security. Yuty's journey ahead entails harnessing its AI-driven edge while navigating these challenges with strategic agility.

In sum, Yuty's voyage towards market dominance, fortified by our AI solution, promises a sustainable combination of personalized excellence and business growth in the industry.

References

Abbas, Q., Fondón, I., & Rashid, M. (2011). Unsupervised skin lesions border detection via two-dimensional image analysis. *Computer Methods and Programs in Biomedicine, 104*(3). https://doi.org/10.1016/j.cmpb.2010.06.016

Abedalla, A., Abdullah, M., Al-Ayyoub, M., & Benkhelifa, E. (2021). Chest X-ray pneumothorax segmentation using U-Net with EfficientNet and ResNet architectures. *PeerJ Computer Science*, *7*, e607.

Atika, L., Nurmaini, S., Partan, R. U., & Sukandi, E. (2022). Image Segmentation for Mitral Regurgitation with Convolutional Neural Network Based on UNet, Resnet, Vnet, FractalNet and SegNet: A Preliminary Study. *Big Data Cognitive Computing*, *6*(4).

Atliha, V., & Šešok, D. (2020). Comparison of VGG and ResNet used as Encoders for Image Captioning. *Open Conference of Electrical, Electronic and Information Sciences* (pp. 1-4). Vilnius: IEEE.

The British Beauty Council. (2023, August 1). Yuty AI: "When AI models are built by people who aren't recognised as diverse, how can they serve every customer?" The British Beauty Council. https://britishbeautycouncil.com/yuty-artificial-intelligence/

Dey, S. (2018). *Hands-On Image Processing with Python: Expert techniques for advanced image analysis and effective interpretation of image data*. Packt Publishing Ltd.

Du, G., Cao, X., Liang, J., Chen, X., & Zhan, Y. (2020). Medical Image Segmentation based on U-Net: A Review. Journal of Imaging Science and Technology, 64(2), 20508–20508. https://doi.org/10.2352/j.imagingsci.technol.2020.64.2.020508

Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. Nature, 542(7639), 115–118. https://doi.org/10.1038/nature21056

Fortune Business Insights. (2022). *Skincare Market Growth and Key Industry Developments 2022-2023.*Global: Fortune Business Insights.

Felzenszwalb, P. F., & Huttenlocher, D. P. (2004). Efficient Graph-Based Image Segmentation. International Journal of Computer Vision, 59(2), 167–181. https://doi.org/10.1023/b:visi.0000022288.19776.77

Gosain, A., & Sardana, S. (2017). Handling class imbalance problem using oversampling techniques: A review. *International Conference on Advances in Computing, Communications and Informatics* (pp. 79-85). Udupi: IEEE.

GlobalData. (2022). *United Kingdom (UK) Health and Beauty Market Size, Trends, Consumer Attitudes and Key Players, 2021-2026.* London: GlobalData.

Grand View Research. (2020). *Personalized Skin Care Products Market Worth $48.65 Billion By 2030.* San Francisco: Grand View Research Inc.

Guo, Y., Liu, Y., Georgiou, T., & Lew, M. S. (2018). A review of semantic segmentation using deep neural networks. *International Journal of Multimedia Information Retrieval, 7*(1), 87-93.

Haenssle, H. A., Fink, C., Schneiderbauer, R., Toberer, F., Buhl, T., Blum, A., Kalloo, A., Hassen, A. B. H., Thomas, L., Enk, A., Uhlmann, L., Alt, C., Arenbergerova, M., Bakos, R., Baltzer, A., Bertlich, I., Blum, A., Bokor-Billmann, T., Bowling, J., & Braghiroli, N. (2018). Man against machine: diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists. Annals of Oncology, 29(8), 1836–1842. https://doi.org/10.1093/annonc/mdy166

Hafiz, A. M., & Bhat, G. M. (2020). A survey on instance segmentation: state of the art. *International Journal of Multimedia Information Retrieval, 9*(1), 171-189.

Haralick, R. M., & Shapiro, L. G. (1985). Image segmentation techniques. *Computer Vision, Graphics and Image Processing, 29*(1), 100-132.https://link.springer.com/chapter/10.1007/978-3-319-24574-4_28#citeas

Hasson, Y., Gül Varol, Dimitrios Tzionas, Igor Kalevatykh, Black, M. J., Laptev, I., & Schmid, C. (2019). Learning Joint Reconstruction of Hands and Manipulated Objects. HAL (Le Centre Pour La Communication Scientifique Directe). https://doi.org/10.1109/cvpr.2019.01208

He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2017). Mask R-CNN. ArXiv.org. https://arxiv.org/abs/1703.06870

Hesamian, M. H., Jia, W., He, X., & Kennedy, P. (2019). Deep Learning Techniques for Medical Image Segmentation: Achievements and Challenges. Journal of Digital Imaging, 32(4), 582–596. https://doi.org/10.1007/s10278-019-00227-x

Kalafatis, S. P., Remizova, N., Riley, D., & Singh, J. (2012). The differential impact of brand equity on B2B co-branding. Journal of Business & Industrial Marketing, 27(8), 623–634. https://doi.org/10.1108/08858621211273574

Kim, K., Ko, E., Lee, M., Mattila, P., & Hoon Kim, K. (2014). Fashion collaboration effects on consumer response and customer equity in global luxury and SPA brand marketing. Journal of Global Scholars of Marketing Science, 24(3), 350–364. https://doi.org/10.1080/21639159.2014.913376

Ko, J., & Cheoi, K. J. (2021). Image-processing based facial imperfection region detection and segmentation. *Multimedia Tools and Applications, 80*(1), 34283–34296.

Li, H.-H., Liao, Y.-H., Huang, Y.-N., & Cheng, P.-J. (2020). Based on machine learning for personalized skin care products recommendation engine. *International Symposium on Computer, Consumer and Control (IS3C)* (pp. 460-462). Taichung City: IEEE.

Li, T., Yu, Z., Phung, E., Duke, B., Kezele, I., & Aarabi, P. (2019). Lightweight Real-time Makeup Try-on in Mobile Browsers with Tiny CNN Models for Facial Tracking. ArXiv:1906.02260 [Cs]. https://arxiv.org/abs/1906.02260

Li, Z., Kamnitsas, K., & Glocker, B. (2019). Overfitting of neural nets under class imbalance: Analysis and improvements for segmentation. *Lecture Notes in Computer Science*, 402–410. https://doi.org/10.1007/978-3-030-32248-9_45

Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., van der Laak, J. A. W. M., van Ginneken, B., & Sánchez, C. I. (2017). A Survey on Deep Learning in Medical Image Analysis. Medical Image Analysis, 42, 60–88. https://doi.org/10.1016/j.media.2017.07.005

Liu, L., Mou, L., Zhu, X. X., & Mandal, M. (2020). Automatic skin lesion classification based on mid-level feature learning. *Computerized Medical Imaging and Graphics*, *84*, 1-11.

Long, J., Shelhamer, E., & Darrell, T. (2015). Fully Convolutional Networks for Semantic Segmentation. (pp. 3431-3440). IEEE Conference on Computer Vision and Pattern Recognition.

Long, J., Shelhamer, E., & Darrell, T. (2015). Fully Convolutional Networks for Semantic Segmentation. (pp. 3431-3440). IEEE Conference on Computer Vision and Pattern Recognition.

Meena, S. R., Soares, L. P., Grohmann, C. H., van Westen, C., Bhuyan, K., Singh, R. P., Floris, M., & Catani, F. (2022). Landslide detection in the Himalayas using machine learning algorithms and U-Net. Landslides, 19(5), 1209–1229. https://doi.org/10.1007/s10346-022-01861-3

Minaee, S., Boykov, Y., Porikli, F., Plaza, A., Kehtarnavaz, N., & Terzopoulos, D. (2022). Image Segmentation Using Deep Learning: A Survey Publisher: IEEE Cite This PDF. *IEEE Transactions on Pattern Analysis and Machine Intelligence* , *44*(7), 3523 - 3542 .

Mutasa, S., Sun, S., & Ha, R. (2020). Understanding Artificial Intelligence based radiology studies: What is overfitting? *Clinical Imaging*, *65*, 96–99. https://doi.org/10.1016/j.clinimag.2020.04.025

Renjun, X., Junliang, Y., Yi, W., & Mengcheng, S. (2022). Fault Detection Method Based on Improved Faster R-CNN: Take ResNet-50 as an Example. *Geofluids, 2022*, 1-9.

ReporterLinker. (2021). *Global Beauty and Personal Care Market, By Product Type, By Distribution Channel By Region, Competition, Forecast and Opportunities, 2026.* Noida: TechSci Research.

Rice, L., Wong, E., & Kolter, Z. (2020). Overfitting in adversarially robust deep learning. Proceedings of the 37th International Conference on Machine Learning, in Proceedings of Machine Learning Research, 119, 8093-8104. https://proceedings.mlr.press/v119/rice20a.html.

Riddhi Kanetkar. (2021, September 2). Meet Yuty, the beauty business building sustainable AI | Sifted. Sifted; Sifted. https://sifted.eu/articles/yuty-sustainable-beauty

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. Lecture Notes in Computer Science, 9351, 234–241. https://doi.org/10.1007/978-3-319-24574-4_28

Roth, H. R., Lu, L., Lay, N., Harrison, A. P., Farag, A., Sohn, A., & Summers, R. M. (2018). Spatial aggregation of holistically-nested convolutional neural networks for automated pancreas localization and segmentation. *Medical image analysis*, *45*, 94-107.

SailThru. (2020, October 9). *Why Personalization Is Especially Important in the Beauty & Cosmetics Industry*. Retrieved from SailThru: https://www.sailthru.com/marketing-blog/personalization-beauty-industry/

Salehi, S. S. M., Erdogmus, D., & Gholipour, A. (2017). Tversky Loss Function for Image Segmentation Using 3D Fully Convolutional Deep Networks. Machine Learning in Medical Imaging, 379–387. https://doi.org/10.1007/978-3-319-67389-9_44

Selvakumar, K., Varalakshmi, L. M., Gulati, K., Upadhyaya, M., & Varalakshmi, L. M. (2022). Skin disease detection using artificial intelligence. *AIP Conference Proceedings*. AIP Publishing.

Snider, E. J., Hernandez-Torres, S. I., & Hennessey, R. (2023). Using ultrasound image augmentation and ensemble predictions to prevent machine-learning model overfitting. *Diagnostics*, *13*(3), 417. https://doi.org/10.3390/diagnostics13030417

Statista. (2023). *Beauty & Personal Care - United Kingdom*. Retrieved from Statista: https://www.statista.com/outlook/cmo/beauty-personal-care/united-kingdom?currency=GBP

Szeliski, R. (2010). Computer Vision: Algorithms and Applications. https://vim.ustc.edu.cn/_upload/article/files/d4/87/71e9467745a5a7b8e80e94007d1b/4cd69b21-85d3-43ba-9935-fd9ae33da82b.pdf

Wakefield Research. (2018, June 27). Majority of Beauty Consumers Struggle Making Purchases, Would Value Assistance From Virtual Beauty Advisors. Retrieved from Automat: https://automat.ai/resources/majority-of-beauty-consumers-struggle-making-purchases-would-value-assistance-from-virtual-beauty-advisors/

Wang, R., Chen, S., Ji, C., Fan, J., & Li, Y. (2022). Boundary-aware context neural network for medical image segmentation. *Medical Image Analysis, 78*(1), 1-13.

Wang, Z., Lin, J., Wang, R., & Zheng, W. (2019). Data Augmentation is More Important Than Model Architectures for Retinal Vessel Segmentation. *International Congress on Integrative Medicine and health* (pp. 48-52). China: Association for Computing Machinery.

Xu, J., Li, M., & Zhu, Z. (2020). Automatic Data Augmentation for 3D Medical Image Segmentation. *Medical Image Computing and Computer Assisted Intervention* (pp. 378-387). Lima: Springer.

Yang, R., & Buenfeld, N. (2001). Binary segmentation of aggregate in SEM image analysis of concrete. *Cement and Concrete Research, 31*(3), 437-441.

Ying, X. (2019). An overview of overfitting and its solutions. *Journal of Physics: Conference Series*, *1168*, 022022. https://doi.org/10.1088/1742-6596/1168/2/022022

Yoon, H., Kim, S., Lee, J., & Yoo, S. (2023). Deep-Learning-Based Morphological Feature Segmentation for Facial Skin Image Analysis. *Diagnostics, 13*(11), 1-13.

Zhang, C., Benz, P., Argaw, D. M., Lee, S., Kim, J., Rameau, F., . . . Kweon, I. S. (2021). ResNet or DenseNet? Introducing Dense Shortcuts to ResNet. *Winter Conference on Applications of Computer Vision* (pp. 3550-3559). IEEE.

Zhang, J., Xie, T., Yang, C., Song, H., Jiang, Z., Zhou, G., Zhang, D., Feng, H., & Xie, J. (2020). Segmenting Purple Rapeseed Leaves in the Field from UAV RGB Imagery Using Deep Learning as an Auxiliary Means for Nitrogen Stress Detection. Remote Sensing, 12(9), 1403. https://doi.org/10.3390/rs12091403

Zhang, J., Zhang, Y., Jin, Y., Xu, J., & Xu, X. (2023). MDU-Net: Multi-scale densely connected U-Net for Biomedical Image segmentation. *Health Information Science and Systems*, *11*(1). https://doi.org/10.1007/s13755-022-00204-9

Zheng, Q., Purwar, A., Zhao, H., Lim, G. L., Li, L., Behera, D., . . . Kot, A. C. (2022). Automatic Facial Skin Feature Detection for Everyone. *arXiv preprint* . arXiv:Identifier.

Zhu, H., Shi, J., & Wu, J. (2019). Pick-and-learn: Automatic quality evaluation for noisy-labeled image segmentation. *Lecture Notes in Computer Science*, 576–584. https://doi.org/10.1007/978-3-030-32226-7_64

Appendix

Table 1.

Model performance on modified L2 regularization.

Method: L2	Mean IoU	Skin Concern Class									Average
		Redness	WFL	Uneven Skintone	Dark Circles	Clogged Pores	HDM	Blemish	Acne	FBS	
L2=0.001	train	0.61	0.66	0.68	0.44	0.00	0.91	0.00	0.18	0.31	0.42
	val	0.30	0.26	0.41	0.11	0.00	0.00	0.00	0.00	0.02	0.12
L2=0.0001	train	0.84	0.85	0.87	0.73	0.82	0.93	0.80	0.91	0.90	0.85
	val	0.31	0.29	0.43	0.18	0.11	0.00	0.25	0.00	0.03	0.18
L2=0.00001	train	0.85	0.87	0.88	0.79	0.81	0.93	0.80	0.91	0.89	0.86
	val	0.35	0.27	0.42	0.18	0.12	0.00	0.21	0.00	0.06	0.18

Notes: WFL stands for Wrinkle Fine Lines; HDM stands for Hyperpigmentation Dark Marks; FBS stands for Freckle Beauty Spots.

Table 2.

Model performance on the modified Dropout rate.

Method: Dropout	Mean IoU	Skin Concern Class									Average
		Redness	WFL	Uneven Skintone	Dark Circles	Clogged Pores	HDM	Blemish	Acne	FBS	
0.4	train	0.82	0.87	0.86	0.79	0.78	0.93	0.78	0.91	0.90	0.85
	val	0.34	0.27	0.43	0.15	0.11	0.00	0.24	0.00	0.05	0.18
0.5	train	0.83	0.86	0.86	0.79	0.80	0.93	0.82	0.91	0.90	0.86
	val	0.36	0.28	0.41	0.18	0.17	0.00	0.25	0.00	0.08	0.19
0.6	train	0.72	0.83	0.79	0.66	0.82	0.93	0.76	0.90	0.89	0.81
	val	0.32	0.28	0.42	0.16	0.09	0.00	0.20	0.00	0.06	0.17

Notes: WFL stands for Wrinkle Fine Lines; HDM stands for Hyperpigmentation Dark Marks; FBS stands for Freckle Beauty Spots. L2 = 0.0001.

Table 3.

Model performance on modified decoder structure.

Method: Decoder Simplification	Mean IoU	Skin Concern Class									
		Redness	WFL	Uneven Skintone	Dark Circles	Clogged Pores	HDM	Blemish	Acne	FBS	Average
Original	train	0.84	0.85	0.87	0.73	0.82	0.93	0.80	0.91	0.90	0.85
	val	0.31	0.29	0.43	0.18	0.11	0.00	0.25	0.00	0.03	0.18
Simplified	train	0.85	0.85	0.78	0.85	0.81	0.89	0.84	0.92	0.84	0.85
	val	0.41	0.39	0.13	0.48	0.15	0.00	0.18	0.00	0.10	0.20

Notes: WFL stands for Wrinkle Fine Lines; HDM stands for Hyperpigmentation Dark Marks; FBS stands for Freckle Beauty Spots. L2 = 0.0001.

Table 4.

Model performance on modified thresholds.

modified thresholds	Mean IoU	Skin Concern Class									
		Redness	WFL	Uneven Skintone	Dark Circles	Clogged Pores	HDM	Blemish	Acne	FBS	Average
0.3	val	0.36	0.37	0.19	0.50	0.15	0.00	0.24	0.00	0.12	0.2145
0.5 (Original)	val	0.40	0.37	0.20	0.46	0.17	0.00	0.23	0.00	0.10	0.2150
0.6	val	0.40	0.36	0.19	0.43	0.17	0.00	0.22	0.00	0.09	0.2060
0.7	val	0.37	0.34	0.17	0.38	0.16	0.00	0.19	0.00	0.07	0.1878

Notes: WFL stands for Wrinkle Fine Lines; HDM stands for Hyperpigmentation Dark Marks; FBS stands for Freckle Beauty Spots. L2 = 0.0001

Table 5.

Enhanced Acne results using ensemble learning.

Acne	Original	Ensemble Learning
IoU	0.002	**0.07**
Recall	0.002	**0.27**