# Exploring Solar Granulation: from IMaX/SUNRISE to DKIST

**Reza Mansouri**
Georgia State University
Atlanta, GA
rmansouri1@student.gsu.edu

**Rafal A. Angryk**
Georgia State University
Atlanta, GA
rangryk@gsu.edu

**Kevin P. Reardon**
National Solar Observatory
Boulder, CO
kreardon@nso.edu

## Abstract

Granules are small cellular structures that populate the solar photosphere and are formed by the dynamic behavior of convection cells. This constant motion generates an evolving pattern of diverse granule types and intergranular regions across the photosphere. Understanding this microscale phenomenon by accurately identifying and classifying the underlying structures is crucial to advance the knowledge of the fundamental physical processes driving solar dynamics. In this study, we leverage images from the IMaX instrument on the SUNRISE balloon-borne telescope, along with their corresponding ground truth masks, to conduct a comparative evaluation of various neural semantic segmentation models. Our best-performing methodology achieves an average mIoU of 0.41 and an average dice coefficient of 0.53 among the classes. Furthermore, we applied the best performing model to high-resolution images from the Daniel K. Inouye Solar Telescope (DKIST) telescope, generating preliminary annotations to facilitate future analysis, making this the first application of these techniques to data from DKIST. The source code is publicly available at `github.com/rezmansouri/imax-to-dkist`.

***Keywords—*** Solar granulation, Photosphere, Semantic segmentation, Deep neural networks, U-Net

## Introduction

The solar photosphere in broadband visible light exhibits a granular pattern, a cellular structure that covers most of the photosphere apart from sunspots, which form due to magnetic flux concentrations. Granules are small, bright, dome-shaped structures that span horizontal scales of thousands of kilometers; see Fig. 1. They evolve rapidly on a time-scale of minutes (Nordlund, Stein, and Asplund 2009) and serve as direct evidence of convection in the solar interior, where hot plasma ascends at the center of each granule, cools at the surface, and descends along the darker edges (Stix 2002).

Granules exhibit diverse patterns, each associated with distinct convective and magnetic phenomena occurring in the solar photosphere. Exploding granules, also referred to as *granules with dots*, are larger granules that expand rapidly and exhibit a characteristic dark dot at their center before fragmenting (Hirzberger et al. 1999). These structures are closely associated with small-scale magnetic flux (Malherbe et al. 2017) (Guglielmino et al. 2020). Another notable structure is the relatively darker, narrow intergranular lane that forms around granules as cooler plasma sinks into the solar interior. These lanes are distinct and easily identifiable in photospheric observations (Nordlund, Stein, and Asplund 2009).

The solar photosphere also features bright points, slender magnetic flux tubes within intergranular lanes, observable in specific spectral bands like Fraunhofer's G band. These bright points are indicators of high magnetic field concentrations, contributing to variations in ultraviolet flux over the solar cycle, which may influence Earth's climate (Riethmüller, T. L. et al. 2014). Additionally, granules with arch-like features, linked to vortex flow tubes and linear magnetic fields, provide further evidence of the complex dynamics within the photosphere (Steiner et al. 2010).

In this study, we address the pixel-level classification of solar granular patterns using semantic segmentation techniques from machine learning. Accurate identification of these patterns is essential for advancing our understanding of solar convection dynamics and magnetic flux emergence. Building on prior work, we explore a range of model configurations and loss functions to determine the most effective approach for this task. Additionally, we examine the effect of encoder-side redundancy reduction prior to training. Finally, we apply our best-performing model to high-resolution DKIST images to generate pre-annotations, laying the groundwork for future research in this domain.

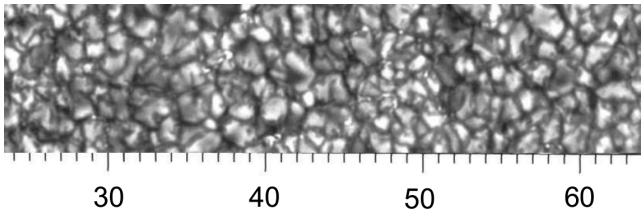


Figure 1: Photosphere granulation, G. Scharmer Swedish Vacuum Solar Telescope 10 July 1997 (Distance in units of 1000 kilometers).

## Related Works

Pixel-level classification is commonly addressed as a supervised learning task in deep learning, specifically semantic segmentation. Fully convolutional neural networks, such as U-Net (Ronneberger, Fischer, and Brox 2015), have demonstrated remarkable performance in this domain. U-Net's encoder-decoder architecture enables systematic feature extraction and upscaling to produce pixel-level class predictions. Variants of U-Net, such as U-Net++ (Zhou

et al. 2018) have been developed to address specific challenges in segmentation tasks. U-Net++ incorporates entangled skip connections with additional convolutional layers, bridging the semantic gap between encoder and decoder sub-networks for improved feature representation. These advancements have proven effective for segmenting complex structures in medical imaging and other domains.

In the context of solar granules, (Díaz Castillo et al. 2022) were the first to address pixel-level classification using deep learning. They employed images from the IMaX instrument on the SUNRISE balloon-borne telescope, annotated with five distinct pattern types, and utilized the original U-Net model. However, their study was limited to a single architecture, leaving room for exploring model variations and optimization techniques.

Challenges such as class imbalance further complicate the task. Class imbalance can bias models toward the majority class, resulting in high overall accuracy but poor performance on minority classes. Addressing these challenges requires careful data sampling, effective preprocessing, appropriate loss functions, and model complexity optimization.

# Methodologies

## Dataset

The granular patterns analyzed in this study are derived from high-resolution observations of the solar photosphere, captured by the Imaging Magnetograph eXperiment (IMaX) (Martínez Pillet et al. 2010) aboard the Sunrise I balloon-borne solar observatory during its June 2009 flight. Our ground-truth dataset consists of eight frames, each with a resolution of 768 × 768 pixels and a field of view of 38" × 38". These spectropolarimetric images correspond to the Fe I 5250.2 Å line, which forms in the lower photosphere and depict the continuum intensity, revealing the granular patterns we aim to classify. Initially, (Díaz Castillo et al. 2022) selected a 56-minute time series with a cadence of 30 seconds, resulting in a total of 113 frames. From these, they identified the eight frames with the highest quality and maximized temporal separation to ensure optimal diversity. To annotate these images, they first executed the MLT4 pattern recognition algorithm (Bovelet and Wiehr 2007) on the images to obtain the underlying pattern and further manually labeled them. These annotated images, along with their established classification scheme, were provided for this study.

As outlined in the introduction, we focus on the same granule categories defined by (Díaz Castillo et al. 2022) for the classification task: granules with dots, granules with lanes, uniform-shaped granules with ellipsoid shapes, complex-shaped granules, and intergranular lane. These five categories were selected as they represent the most significant patterns linked to key physical phenomena occurring in the photosphere. We used 7 frames for training and one specific frame for validation during all our experiments. We observed an extreme class-imbalance in the data, with intergranular lanes and complex-shaped granules being the dominant ones, as shown in Fig. 2.

Using these statistics, we assign importance weights to each of our classes, being equal to the inverse of their frequency. To create our training and validation sets, each frame is first augmented with all the rotations possible in increments of 5 degrees. Then we perform random sub-patch cropping of size 128 × 128 pixels, with the probability of each pixel being the center of our selection as its class's importance weight, while ignoring the invalid selections in the edges of the image. This random sampling method represents our initial approach to addressing the class imbalance issue, aiming to create a dataset with a uniform distribution across the classes. Afterward, random flipping both vertically and horizontally is applied. Using this approach, we generated a training set of size 28500, and a validation set of 3750. See some instances in Fig. 3.

## Model architectures

**U-Net** Initially proposed and used in medical imaging, the U-Net (Ronneberger, Fischer, and Brox 2015) is the most common model in semantic segmentation. Its architecture (Fig. 4) features an encoder-decoder structure, where the encoder progressively reduces the spatial dimensions of the input image while extracting relevant features, and the decoder reconstructs the image to the original size, facilitating the generation of segmented outputs. This unique design incorporates skip connections, allowing high-resolution features from the encoder to be combined with the upsampled outputs in the decoder. This helps preserve spatial information and enhances the model's performance. In our experiments, each convolutional block in the encoding subnet of the U-Net consists of two consecutive repetitions of a 3×3 convolution, followed by batch normalization and a ReLU activation function. To maintain the dimensions of the data throughout this pipeline, we employed a padding of 1 in each convolutional layer. The output from this block is then subjected to 2×2 max pooling, which reduces the spatial dimensions while enabling the extraction of higher-level features. In the decoder subnet, each convolutional block receives the output from its corresponding encoding block and concatenates it with the output of the 2×2 transposed convolution from the block below. This process effectively upsamples the features back to a higher dimension. It should also be noted that to mitigate overfitting, channel-wise dropout with a probability of 0.5 is applied after each max-pooling operation and following each upsampling step via transposed convolution. Finally, a 1×1 convolution is applied to match the number of output channels to the number of classes for prediction, followed by a softmax activation along the channels to get the final prediction scores for each pixel. This structured approach facilitates effective feature learning and enhances the model's segmentation capabilities.

**U-Net++** U-Net++ (Zhou et al. 2018) is an improved version of the original U-Net architecture, designed to enhance image segmentation tasks. U-Net++ adds dense skip pathways (Fig. 5) that improve feature sharing between the encoder and decoder. This allows the network to capture more detailed information at different levels, resulting in better segmentation accuracy. While it main-

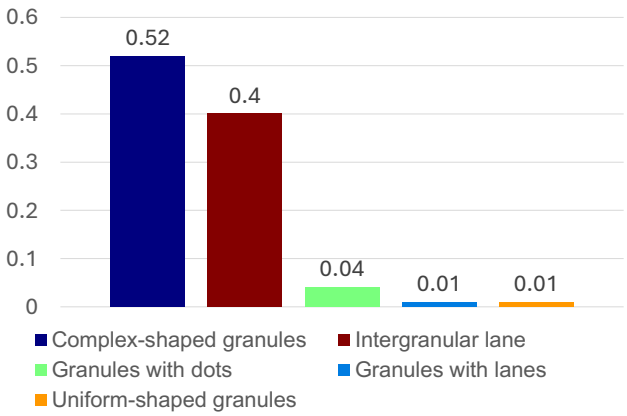


Figure 2: Distribution of granular patterns in the IMaX dataset

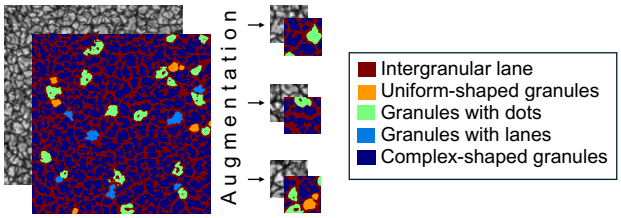


Figure 3: Some instances from augmenting IMaX data

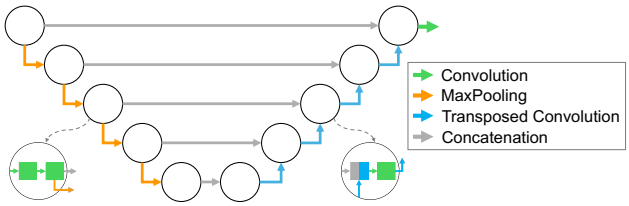


Figure 4: The U-Net architecture

tains the encoder-decoder structure of U-Net, the additional convolutional blocks in the skip connections help create richer feature representations. One of the key benefits of U-Net++ is its ability to address the vanishing gradient problem in deeper networks. The multiple skip pathways allow the model to effectively combine low-level and high-level features, making it more robust against overfitting. Similar to our choices for U-Net, we implemented similar combinations of layers within the convolutional blocks of both the encoder and decoder subnets in U-Net++. The architecture retains the same fundamental structure, ensuring that the convolutional blocks in the middle are consistent with those in the decoder.

## Model complexity

Given that this dataset is new and lacks established benchmarks, the extent of high-level features within the granules remains uncertain. To explore this, we employed various combinations of models by adjusting the number of levels in both the encoder and decoder subnets, as well as experimenting with different number of kernels. Specifically, we tested U-Net and U-Net++ architectures with 3, 4, and 5 levels. The 5-level models correspond to the originally proposed architectures, while the simpler configurations are classified as pruned models (Fig. 6). For the starting kernels, we utilized 16, 32, and 64, with the number of kernels doubling as we progressed down each block, resulting in configurations of (16, 32, 64), (32, 64, 128), and (64, 128, 256) for a 3-level model. This structure was maintained across the 4-level and

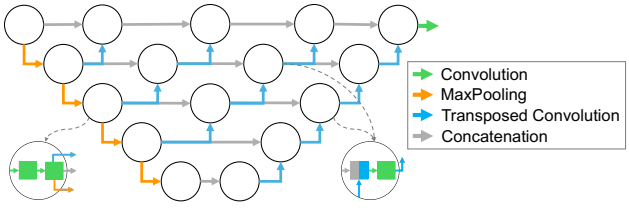


Figure 5: The U-Net++ architecture

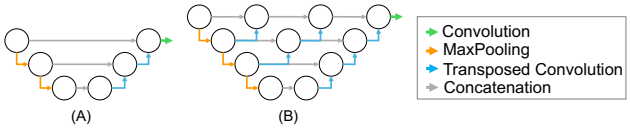


Figure 6: Pruned architectures of (A) U-Net and (B) U-Net++

5-level architectures, yielding a total of nine distinct architectures for both U-Net and U-Net++. Through this approach, we aimed to uncover the underlying complexity of the data while also striking a reasonable balance between model complexity and accuracy. We will denote a U-Net or U-Net++ architecture with a complexity of $(k, k \times 2, k \times 4, \ldots, k \times 2^{n-1})$ as $k..k \times 2^{n-1}$, where $k$ is the starting number of kernels and $n$ is the number of levels in the encoder/decoder subnets.

## Loss functions

As we are facing a multi-class classification problem from a simple perspective, cross-entropy loss is commonly used. However, it falls short in adapting to our class imbalance, making it necessary to employ more sophisticated methods to tackle the unique challenges posed by our dataset. In this section we cover our candidates of loss functions to be optimized.

**Mean IoU** Intersection over Union (IoU) is a widely used evaluation metric for image segmentation models, also known as the Jaccard index in basic classification tasks (Fig. 7). It quantifies the overlap between the predicted segmentation and the ground truth by calculating the intersection over the union of the predicted and actual regions. To obtain an aggregated result across all classes, we compute the *mean IoU* (mIoU) by averaging the IoU over $N$ classes:

$$\text{mIoU} = \frac{1}{N} \sum_{i=1}^{N} \frac{TP_i}{TP_i + FP_i + FN_i} \tag{1}$$

where $TP_i$, $FP_i$, and $FN_i$ denote the true positives, false positives, and false negatives for class $i$, respectively.

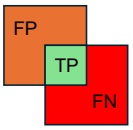


Figure 7: The analogy between the Jaccard index and IoU for image segmentation, with the orange square as the predicted object, and the red one as the ground truth

mIoU is a similarity measure. As a loss function needs to be minimized, we need to transfer mIoU to a dissimilarity function of the predictions and the ground truth. It can be done by subtracting it from 1. Moreover, in addition to the sampling method used for data augmentation, we also applied a weighted loss function to address class imbalance. By weighting the loss based on each class's distribution, we ensure that the contribution of each class is considered, especially in cases of incorrect classification.

The formula is:

$$\mathcal{L}_{\text{mIoU}} = 1 - \frac{1}{N} \sum_{i=1}^{N} \frac{p_i}{TP_i + FP_i + FN_i} \times w_i \tag{2}$$

With $p_i$ being the predicted probability of a pixel being correctly classified as class $i$, and $w_i$ the inverse of class $i$'s frequency.

**Lovász-Softmax** Derived from the mIoU, this loss function (Berman, Triki, and Blaschko 2018) is specifically crafted for direct optimization in neural networks, utilizing the convex Lovász extension of submodular losses. The loss function directly optimizes the ranked errors based on their impact on the Jaccard index. Misclassified pixels that have a higher impact on the overall IoU (because of their large intersection or union contribution) are weighted more, allowing the network to prioritize these pixels during optimization. This loss function is defined as:

$$\mathcal{L}_{\text{Lovász-Softmax}}(y, \hat{y}) = \frac{1}{N} \sum_{i=1}^{N} \mathcal{L}_{\text{Lovász}}(m^{(i)}) \tag{3}$$

- $N$: Number of classes.
- $\mathcal{L}_{\text{Lovász}}(m^{(i)})$: Lovász Softmax Loss for class $i$.
- $m^{(i)}$: Classification error for class $i$, defined as:

$$m^{(i)} = \begin{cases} 1 - \hat{y} & \text{if } y = 1, \\ \hat{y} & \text{if } y = 0, \end{cases}$$

  where $y$ is the ground truth label and $\hat{y}$ is the predicted probability for belonging to class $i$.
- $\mathcal{L}_{\text{Lovász}}$: Calculated as:

$$\mathcal{L}_{\text{Lovász}}(m^{(i)}) = m^{(i)} \Delta^{(i)}$$

  where $m^{(i)}$ are the sorted errors in descending order for class $i$, and $\Delta^{(i)}$ is the gradient of the Jaccard index.

It is important to note that the definitions of both of our loss functions have been simplified for the classification of a single pixel. In practice, we will mean-aggregate the calculations across all dimensions for acquiring the losses and performance metrics.

## Redundancy reduction

BT-UNet is a self-supervised learning framework (Punn and Agarwal 2022) originally proposed for biomedical image segmentation, leveraging the redundancy reduction strategy known as Barlow Twins (Zbontar et al. 2021) for pre-training the encoder component of the U-Net in an unsupervised manner. This framework aims to enhance U-Net's performance while addressing the challenge of limited annotated data, a common issue we encountered. To achieve accurate and redundancy-free representation learning, we begin by augmenting our dataset of 128 × 128 images with various distortions, such as random solarization and rotations for each instance. This process generates a corresponding set of corrupted images for the pre-training step. Subsequently, the encoder of the U-Net, connected to several fully connected layers known as the projection network, is trained using a Siamese network approach (Taigman et al. 2014) that involves two forward propagations: one on the first distorted image and another on the second. This architecture creates two dense feature representations, as illustrated in Fig. 8. The batch-normalized representations are then utilized to construct a cross-correlation matrix, which the loss function (Eq. 4) aims to make as close to the identity matrix as possible. The diagonal term $c_{ii}$ encourages the model to produce similar outputs for corresponding features in both representations, while the off-diagonal term $c_{ij}$ promotes different outputs across distinct features, guiding the model toward redundancy reduction. The term $\lambda$ balances the contributions of the diagonal and off-diagonal terms in the loss function, and we set it to 0.2 for our experiments.

$$\mathcal{L}_{\text{Barlow}} = \sum_{i=1}^{N} (1 - c_{ii})^2 + \lambda \sum_{i \neq j} c_{ij}^2 \tag{4}$$

## Training and inference

We report our results using three main metrics: accuracy, IoU, and the Dice coefficient (also referred to as the F1 score) for each class. Additionally, we compute average metrics across all classes to provide an initial evaluation of our models' overall performance.

$$\text{Dice} = \frac{2 \cdot TP}{2 \cdot TP + FP + FN} \tag{5}$$

For the final prediction of a single 768 × 768 frame, one approach is to divide it into 36 non-overlapping sub-patches of 128 × 128 pixels, and then stitch the resulting segmentation sub-masks. However, this method is prone to failure due to potential inconsistencies in classifying granular artifacts—since the spatial context between adjacent sub-patches is lost. To address this issue, we instead used a 32-strided overlapping prediction window, accumulating the softmax prediction scores across all overlapping passes. After completing the sweep, we apply an `argmax` operation on the aggregated score map at each pixel to assign the final class label. This approach ensures that the segmentation mask preserves the spatial continuity of the entire frame.

For training our architectures, we used a mini batch size of 64 and performed each experiment for 100 epochs. Our employed models are the ones with the least validation loss during training. We used the Adam optimizer (Kingma and Ba 2017), with a starting learning rate of $10^{-3}$. For pre-training the encoder subnets of our models, we performed training for 50 epochs and used adam with learning rate of $10^{-6}$. For both training and pre-training, every time the minimization of the validation loss failed to improve for 5 consecutive epochs, the learning rate was adjusted to $0.9\times$ its previous value. We used PyTorch (Paszke et al. 2019) version 2.2.1 in Python 3.10.12. The experiments were run on an NVIDIA A40 48Gb GPU, 22.04.3 Ubuntu system with ∼500Gb memory.

## Prediction on DKIST images

The National Science Foundation's (NSF) Daniel K. Inouye Solar Telescope (DKIST) is currently the largest solar telescope in the world, boasting a four-meter aperture. Notably, the quality of DKIST images surpasses that of IMaX, revealing distinct artifacts such as bright points visible in Fig. 10 (B). On May 26, 2022, DKIST captured a time series of 200 frames at a 6-second cadence. From this dataset, we selected the 10 highest-quality frames that were temporally well-distributed.

After identifying the best model based on IMaX data, we established a procedure to generate initial annotations for these 10 frames. This step is essential for initiating future research involving a supervised-learning solution, with plans for further manual refinement of the labels. Our goal is to ensure that the instances in DKIST closely resemble those in IMaX, allowing the patterns learned by our best models to be effectively transferred to the high-quality DKIST data.

The original DKIST images cover a field of view (FOV) of 45" × 45" with a resolution of 4096 × 4096 pixels. To match the FOV of IMaX, we first center-cropped the DKIST images to 3454 × 3454 pixels. For reference in subsequent transformations, we stitched together all the training frames from IMaX. We then performed histogram matching between each DKIST instance and the IMaX reference to ensure consistency, followed by Gaussian blurring to achieve a clarity that aligns with IMaX images.

Finally, we resized the transformed frames to 768 × 768 pixels, preparing a dataset that can be effectively utilized with any model trained on IMaX. Consistent with the overlapping 128 × 128 windows approach used for IMaX, we applied a stride of 32 during the prediction of preliminary annotations (Fig. 9). This technique helps maintain spatial context between the cropped sub-patches, ensuring

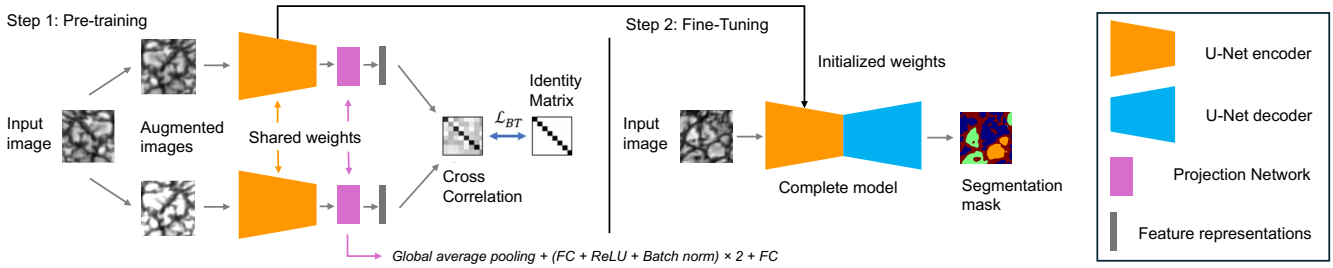


Figure 8: The BT-UNet redundancy reduction framework

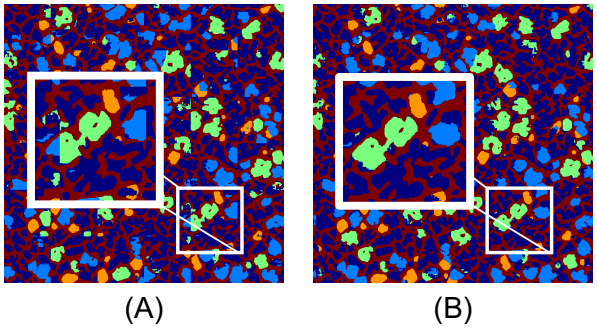


Figure 9: Predicted segmentation mask for a training frame of IMaX using UNet++ 64..512 and mIoU loss: (A) Non-overlapping prediction windows, (B) Overlapping prediction windows with 32 pixel strides

that inconsistent classification for a single object becomes unlikely. Once we obtained the segmentation masks, we resized them back to 3454 × 3454 pixels using nearest-neighbor interpolation, resulting in the creation of final *pre-annotations* (Fig. 10).

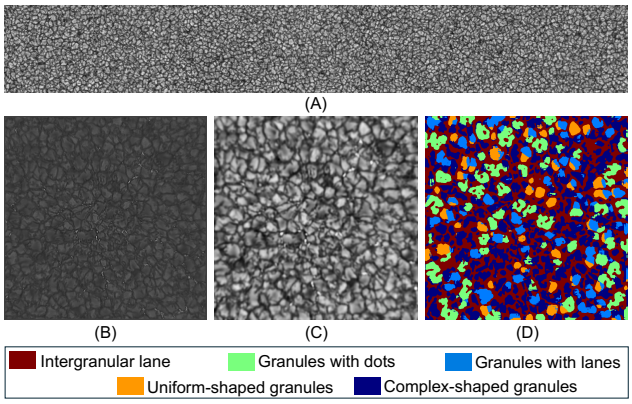


Figure 10: Initial annotations for a DKIST instance (B), transformed instance (C), and predicted segmentation mask (D) using U-Net++ 64..512 trained on IMaX with mIoU loss. (A) is the concatenation of IMaX training frames as reference for transformations.

## Results and Discussion

With the combination of two architectures, U-Net and U-Net++, and two distinct loss functions, we experimented systematically on the IMaX dataset. To evaluate the impact of redundancy reduction, we included the BT-UNet variant and explored nine possible model complexities.

Given that average metrics are often skewed towards dominant classes, our initial evaluations focused on average scores across all classes. However, to gain a more granular understanding, we planned to further filter these results based on the model's performance on individual classes. The reported outcomes correspond to the validation frame, providing a comprehensive assessment of each model's behavior.

To begin our analysis, we will investigate the effects of model complexity and our chosen loss functions. In our initial set of experiments, we trained U-Net architectures across all complexities using the mIoU loss function. As shown in Table 1, the architecture with 64..512 complexity yielded the best average results per class. This finding highlights how well granular features are extracted, which is optimally achieved through four levels of feature extraction, starting with a base of 64 convolutional kernels.

Moving on to the second set of experiments, we conducted a similar test using the Lovász-Softmax loss function, as shown in Table 1. The 64..512 architecture once again demonstrated its superiority. Although the results are comparable to those obtained with the mIoU loss function, they show slightly poorer performance. This observation suggests that the Lovász-Softmax loss function may not be the most suitable choice for optimizing our model, considering the characteristics of our data and the challenges associated with the multi-class segmentation task at hand.

In the third set of experiments, we explored the pre-training strategy using BT-UNet. We selected the three most complex architectures, which had previously shown promising results, and investigated both loss functions, as presented in Table 1.

The findings indicate that models trained with pre-training tend to achieve better results on average across different architectures. This improvement can be attributed to the model learning relatively general and accessible features during the pre-training phase. Consequently, the main training benefits from a more favorable initialization, enabling the decoder sub-network to more effectively reconstruct features to match the original image dimensions.

Finally, we examined the U-Net++ architecture, following a similar methodology as in the first two sets of experiments. The results are presented in Table 1. As shown, U-Net++ consistently outperforms our previous models, demonstrating its effectiveness in handling the semantic segmentation task. Notably, the mIoU loss function once again yields better results than the Lovász-Softmax loss, reinforcing the advantages of using mIoU in our context. We did not include the 64..1024 configuration in these experiments, as it exhibited clear signs of overfitting in both U-Net and BT-UNet,

Table 1: Average per class results for U-Net, BT-UNet, and U-Net++

| Complexity | U-Net | | | | | | BT-UNet | | | | | | U-Net++ | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | mIoU Loss | | | Lovász Softmax Loss | | | mIoU Loss | | | Lovász Softmax Loss | | | mIoU Loss | | | Lovász Softmax Loss | | |
| | **Acc.** | **mIoU** | **Dice** | **Acc.** | **mIoU** | **Dice** | **Acc.** | **mIoU** | **Dice** | **Acc.** | **mIoU** | **Dice** | **Acc.** | **mIoU** | **Dice** | **Acc.** | **mIoU** | **Dice** |
| 16..64 | 0.43 | 0.29 | 0.38 | 0.41 | 0.20 | 0.25 | | | | | | | 0.50 | 0.33 | 0.44 | 0.52 | 0.32 | 0.43 |
| 16..128 | 0.46 | 0.35 | 0.45 | 0.50 | 0.31 | 0.40 | | | | | | | 0.49 | 0.35 | 0.46 | 0.50 | 0.38 | 0.48 |
| 16..256 | 0.44 | 0.24 | 0.33 | 0.43 | 0.22 | 0.28 | | | | | | | 0.56 | 0.41 | 0.53 | 0.44 | 0.32 | 0.40 |
| 32..128 | 0.50 | 0.35 | 0.46 | 0.45 | 0.31 | 0.38 | | | | | | | 0.50 | 0.37 | 0.47 | 0.49 | 0.30 | 0.40 |
| 32..256 | 0.53 | 0.40 | 0.51 | 0.52 | 0.38 | 0.48 | | | | | | | 0.55 | 0.39 | 0.50 | 0.48 | 0.39 | 0.49 |
| 32..512 | 0.47 | 0.33 | 0.43 | 0.48 | 0.32 | 0.42 | | | | | | | 0.55 | 0.39 | 0.51 | 0.48 | 0.33 | 0.43 |
| 64..256 | 0.48 | 0.34 | 0.44 | 0.48 | 0.30 | 0.38 | 0.49 | 0.37 | 0.48 | 0.50 | 0.34 | 0.44 | 0.51 | 0.37 | 0.47 | 0.47 | 0.28 | 0.38 |
| 64..512 | 0.57 | 0.37 | 0.50 | 0.53 | 0.37 | 0.48 | 0.55 | 0.42 | 0.54 | 0.53 | 0.36 | 0.47 | 0.58 | 0.41 | 0.53 | 0.50 | 0.30 | 0.39 |
| 64..1024 | 0.50 | 0.39 | 0.50 | 0.52 | 0.36 | 0.46 | 0.51 | 0.38 | 0.49 | 0.54 | 0.35 | 0.46 | | | | | | |

Table 2: Class-specific results for models with 64..512 complexity and mIoU loss

| Model | CG | | | IL | | | GD | | | GL | | | UG | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Acc. | IoU | Dice coeff. | Acc. | IoU | Dice coeff. | Acc. | IoU | Dice coeff. | Acc. | IoU | Dice coeff. | Acc. | IoU | Dice coeff. |
| U-Net | 0.58 | 0.49 | 0.66 | 0.86 | 0.82 | 0.90 | 0.48 | 0.29 | 0.46 | 0.58 | 0.08 | 0.15 | 0.33 | 0.19 | 0.31 |
| BT-UNet | 0.72 | 0.61 | 0.75 | 0.89 | 0.84 | 0.91 | 0.42 | 0.31 | 0.47 | 0.29 | 0.06 | 0.11 | 0.43 | 0.29 | 0.45 |
| U-Net++ | 0.66 | 0.58 | 0.73 | 0.94 | 0.86 | 0.92 | 0.46 | 0.33 | 0.50 | 0.46 | 0.09 | 0.16 | 0.37 | 0.22 | 0.36 |

**CG**: Complex-shaped granules, **IL**: Intergranular lane, **GD**: Granules with dots, **GL**: Granules with lanes, **UG**: Uniform-shaped granules

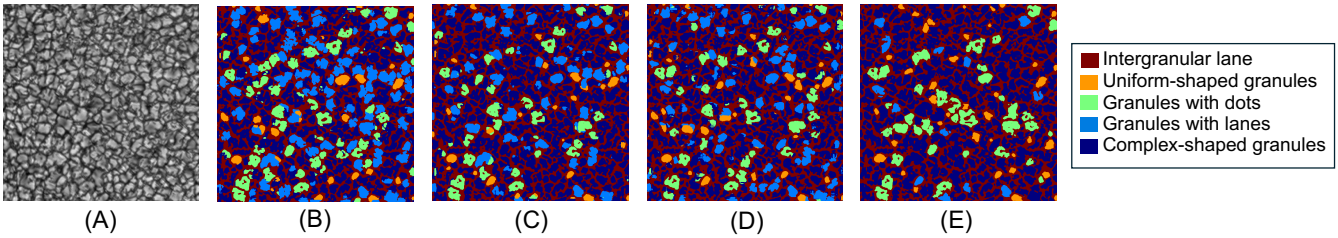


Figure 11: Final predicted segmenantion masks on (A) the validation frame using (B) U-Net, (C) BT-UNet, and (D) UNet++, with 64..512 architecture and trained with mIoU loss, and (E) ground truth segmentation mask

making it a less suitable candidate for further evaluation.

The success of U-Net++ can be attributed to its unique design, particularly the use of skip connections intertwined with convolutional blocks in the middle layers. These skip connections play a crucial role in bridging the gap between the feature maps generated by the encoder and those produced by the decoder. By allowing features to flow more freely between these layers, the model can better capture and represent complex details in the data. This leads to a more accurate representation of features, ultimately enhancing the model's performance in semantic segmentation tasks.

Table 2 shows the detailed results for each class, allowing for a clear comparison of models perform across different classes, for U-Net, BT-UNet, and U-Net++ using our best-performing complexity of 64..512 and the mIoU loss function. Corresponding visualizations are presented in Fig. 11.

## Conclusion

In summary, this study presents a comprehensive exploration of the granular structures in the solar photosphere through advanced neural semantic segmentation techniques. Utilizing high-resolution data from the IMaX instrument aboard the SUNRISE telescope and applying our methods to DKIST imagery, we enhance the identification and classification of granules and intergranular regions, contributing to a better understanding of their dynamic morphology. The preliminary annotations generated for DKIST images provide a solid foundation for future investigations, contributing to a deeper understanding of solar dynamics. This research not only underscores the potential of deep learning approaches in solar physics but also paves the way for further studies aimed at unraveling the fundamental processes that govern the Sun's behavior. Future work should explore the capabilities of other more advanced models inspired by the U-Net architecture to achieve even greater accuracy and insight.

## References

[Berman, Triki, and Blaschko 2018] Berman, M.; Triki, A. R.; and Blaschko, M. B. 2018. The lovász-softmax loss: A tractable surrogate for the optimization of the intersection-over-union measure in neural networks.

[Bovelet and Wiehr 2007] Bovelet, B., and Wiehr, E. 2007. Multiple-scale pattern recognition applied to faint intergranular g-band structures. *Solar Physics* 243(2):121–129.

[Díaz Castillo et al. 2022] Díaz Castillo, S.; Asensio Ramos, A.; Fischer, C.; and Berdyugina, S. 2022. Towards the identification and classification of solar granulation structures using semantic segmentation. *Frontiers in Astronomy and Space Sciences* 9:896632.

[Guglielmino et al. 2020] Guglielmino, S. L.; Pillet, V. M.; Cobo, B. R.; Rubio, L. R. B.; del Toro Iniesta, J. C.; Solanki, S. K.; Riethmüller, T. L.; and Zuccarello, F. 2020. On the magnetic nature of an exploding granule as revealed by sunrise/imax. *The Astrophysical Journal* 896(1):62.

[Hirzberger et al. 1999] Hirzberger, J.; Bonet, J.; Vázquez, M.; and Hanslmeier, A. 1999. Time series of solar granulation images.

iii. dynamics of exploding granules and related phenomena. *The Astrophysical Journal* 527:405–414.

[Kingma and Ba 2017] Kingma, D. P., and Ba, J. 2017. Adam: A method for stochastic optimization.

[Malherbe et al. 2017] Malherbe, J.-M.; Roudier, T.; Stein, R.; and Frank, Z. 2017. Dynamics of trees of fragmenting granules in the quiet sun: Hinode/sot observations compared to numerical simulation. *Solar Physics* 293(1).

[Martínez Pillet et al. 2010] Martínez Pillet, V.; del Toro Iniesta, J. C.; Álvarez Herrero, A.; Domingo, V.; Bonet, J. A.; González Fernández, L.; López Jiménez, A.; Pastor, C.; Gasent Blesa, J. L.; Mellado, P.; Piqueras, J.; Aparicio, B.; Balaguer, M.; Ballesteros, E.; Belenguer, T.; Bellot Rubio, L. R.; Berkefeld, T.; Collados, M.; Deutsch, W.; Feller, A.; Girela, F.; Grauf, B.; Heredero, R. L.; Herranz, M.; Jerónimo, J. M.; Laguna, H.; Meller, R.; Menéndez, M.; Morales, R.; Orozco Suárez, D.; Ramos, G.; Reina, M.; Ramos, J. L.; Rodríguez, P.; Sánchez, A.; Uribe-Patarroyo, N.; Barthol, P.; Gandorfer, A.; Knoelker, M.; Schmidt, W.; Solanki, S. K.; and Vargas Domínguez, S. 2010. The imaging magnetograph experiment (imax) for the sunrise balloon-borne solar observatory. *Solar Physics* 268(1):57–102.

[Nordlund, Stein, and Asplund 2009] Nordlund, Å.; Stein, R. F.; and Asplund, M. 2009. Solar surface convection. *Living Reviews in Solar Physics* 6(1):1–117.

[Paszke et al. 2019] Paszke, A.; Gross, S.; Massa, F.; Lerer, A.; Bradbury, J.; Chanan, G.; Killeen, T.; Lin, Z.; Gimelshein, N.; Antiga, L.; Desmaison, A.; Köpf, A.; Yang, E.; DeVito, Z.; Raison, M.; Tejani, A.; Chilamkurthy, S.; Steiner, B.; Fang, L.; Bai, J.; and Chintala, S. 2019. Pytorch: An imperative style, high-performance deep learning library.

[Punn and Agarwal 2022] Punn, N. S., and Agarwal, S. 2022. Bt-unet: A self-supervised learning framework for biomedical image segmentation using barlow twins with u-net models.

[Riethmüller, T. L. et al. 2014] Riethmüller, T. L.; Solanki, S. K.; Berdyugina, S. V.; Schüssler, M.; Martínez Pillet, V.; Feller, A.; Gandorfer, A.; and Hirzberger, J. 2014. Comparison of solar photospheric bright points between sunrise observations and mhd simulations. *AA* 568:A13.

[Ronneberger, Fischer, and Brox 2015] Ronneberger, O.; Fischer, P.; and Brox, T. 2015. U-net: Convolutional networks for biomedical image segmentation.

[Steiner et al. 2010] Steiner, O.; Franz, M.; González, N. B.; Nutto, C.; Rezaei, R.; Pillet, V. M.; Navarro, J. A. B.; del Toro Iniesta, J. C.; Domingo, V.; Solanki, S. K.; Knölker, M.; Schmidt, W.; Barthol, P.; and Gandorfer, A. 2010. Detection of vortex tubes in solar granulation from observations with sunrise. *The Astrophysical Journal Letters* 723(2):L180.

[Stix 2002] Stix, M. 2002. *The Sun: An Introduction*. Germany: Springer.

[Taigman et al. 2014] Taigman, Y.; Yang, M.; Ranzato, M.; and Wolf, L. 2014. Deepface: Closing the gap to human-level performance in face verification. In *2014 IEEE Conference on Computer Vision and Pattern Recognition*, 1701–1708.

[Zbontar et al. 2021] Zbontar, J.; Jing, L.; Misra, I.; LeCun, Y.; and Deny, S. 2021. Barlow twins: Self-supervised learning via redundancy reduction.

[Zhou et al. 2018] Zhou, Z.; Siddiquee, M. M. R.; Tajbakhsh, N.; and Liang, J. 2018. Unet++: A nested u-net architecture for medical image segmentation.