Journal of Food Engineering 243 (2019) 9–21

Contents lists available at ScienceDirect

Journal of Food Engineering

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

Visual modeling of laser-induced dough browning

Peter Yichen Chen[a,*], Jonathan David Blutinger[a], Yorán Meijers[a,b], Changxi Zheng[a], Eitan Grinspun[a], Hod Lipson[a]

[a] Columbia University, 116th St & Broadway, New York, NY, 10027, USA
[b] Wageningen University, 6708 PB Wageningen, The Netherlands

ARTICLE INFO

Keywords:
Infrared laser
Dough
Browning
Generative model
Deep learning
Deconvolution

ABSTRACT

A data-driven model that predicatively generates photorealistic RGB images of dough surface browning is proposed. This model was validated in a practical application using a CO_2 laser dough browning pipeline, thus confirming that it can be employed to characterize visual appearance of browned samples, such as surface color and patterns. A supervised deep generative network takes laser speed, laser energy flux, and dough moisture as an input and outputs an image (of 64 × 64 pixel size) of laser-browned dough. Image generation is achieved by nonlinearly interpolating high-dimensional training data. The proposed prediction framework contributes to the development of computer-aided design (CAD) software for food processing techniques by creating more accurate photorealistic models.

1. Introduction

Visualization tools—what designers use to communicate complex concepts through rendered images—are of preeminent importance when new food products are designed. Modeling software that can produce more accurate visuals of processed food will aid in the development of this technology by creating more compelling and realistic models. The aim of this paper is to create a robust model for generating laser-induced dough browning.

Products that undergo Maillard browning (Martins et al., 2000), such as baked bread, demonstrate rich surface color and pattern, which are key characteristics for quality control and consumer appeal (Purlis and Salvadori, 2007). In extant studies, the surface color of browned products is typically modeled by applying (1) a physics-based method (Purlis and Salvadori, 2009; Zanoni et al., 1995; Zhang and Datta, 2006) or (2) a data-driven (machine learning, observation-based) approach (Purlis and Salvadori, 2007).

When the physics-based method is adopted, surface color is usually characterized using scalar values, such as browning index and surface lightness. First-order kinetics are subsequently employed to model this scalar value as a function of bread surface temperature. Coupled heat and mass transfer equations are solved numerically to calculate surface temperature, which serves as the browning index indicator. This method is not only computationally expensive, but also requires tuning of a large number of parameters (Zhang and Datta, 2006).

Purlis and Salvadori (2007) took a data-driven approach to characterize surface color of bread browning as a function of baking temperature. The authors established a linear relationship between total color change and weight based on the best fit to the experimentally obtained data. Physics-based and data-driven methods have both been shown to qualitatively predict a browning behavior trend, both spatially and temporally, albeit with limited accuracy (Purlis and Salvadori, 2007; Zanoni et al., 1995). In particular, no attempt has been made to obtain full surface color (either RGB or $L^*a^*b^*$, i.e., not just a scalar value) using either of these methods.

Contrary to the abundant research pertaining to dough surface color modeling, surface pattern modeling has been a subject of limited studies. Purlis and Salvadori (2009) observed that proofed dough under browning results in a wrinkled surface, which becomes smooth after a few minutes. However, no attempt was made by these authors to elucidate the relationship between the browned product smoothness and the baking setup or dough recipe.

In this work, dough surface pattern modeling was driven by an empirical observation that surface pattern is automatically captured if surface color is modeled in a full color space, instead of the over-simplified scalar browning index or brightness. In other words, if the images generated by a computational model can be as realistic as those captured by camera, pattern information can be easily obtained from the generated images.

While surface color in full color space can be modeled using a

* Corresponding author.
E-mail addresses: cyc@cs.columbia.edu (P.Y. Chen), jdb2202@columbia.edu (J.D. Blutinger), ymeijers@gmail.com (Y. Meijers), cxz@cs.columbia.edu (C. Zheng), eitan@cs.columbia.edu (E. Grinspun), hod.lipson@columbia.edu (H. Lipson).

https://doi.org/10.1016/j.jfoodeng.2018.08.022
Received 18 April 2018; Received in revised form 20 August 2018; Accepted 21 August 2018
Available online 25 August 2018
0260-8774/

physics-based method, generating photorealistic color using this method is challenging due to the complexity of the browning phenomenon (Martins et al., 2000; Purlis and Salvadori, 2009). It is likely that, owing to these difficulties, the authors of the existing surface color models have resorted to color modeling by simply linearizing the relationship between color and temperature and using only a scalar value for color description.

Data-driven methods overcome the need to model intermediate physical and chemical processes (Sablani, 2008) because they are applied directly to input and output data. This is particularly useful when the underlying physical process is not well understood, as is the case with browning. For example, when browning models are based on a data-driven method, they can be applied directly on photographs of the browned sample to construct a correspondence between the output photograph and the input comprising of browning apparatus parameters.

Data-driven methods are also much faster compared to physics-based methods as, once the machine learning model is trained, it generates a prediction in real time with low computational cost. On the contrary, whenever a different material property or boundary condition is needed, physics-based methods have to re-solve multiple governing differential equations using numerical approximations. The computational cost can thus be prohibitive, as the processing can take minutes, hours, and even days (Purlis and Salvadori, 2009; Zhang and Datta, 2006).

Data-driven methods have limitations as well. Because the underlying physical and chemical processes are unknown, this method is often considered a black box, the accuracy of which can only be tested empirically. Since data-driven models are trained on a specific dataset, attention is needed to prevent overfitting and the ability to extrapolate is limited.

In light of the aforementioned benefits and drawbacks, the aim of the present study was exploring data-driven method's potential to predict high-resolution, full-color, photorealistic images of browned products, which capture both surface color and pattern.

In particular, in this work, a deep learning approach was employed to model photorealistic browned dough. In the past decade, deep learning has become a state-of-the-art machine learning method with extensive AI applications, especially in object recognition (LeCun et al., 2015). In a typical object recognition problem, an RGB image at pixel level would serve as input for classifying different objects in the image. This strategy has been employed for recognizing human faces in photographs (LeCun et al., 2015) and detecting diseased plants in maize field images (DeChant et al., 2017), among other applications.

In food science, deep learning has been used to assess puréed food concentrations (Pfisterer et al., 2018) and to classify food and drink images (Mezgec and Koroušić Seljak, 2017). Shallow artificial neural network, as the predecessor of deep learning, has been widely used in food science for food classification, quality prediction, and segmentation of food images (Du and Sun, 2006).

Deep learning is also suitable for generative modeling (Goodfellow et al., 2016). As shown in Fig. 2, generative modeling solves the inverse problem of object recognition, where the object label is an input and the RGB image at pixel level is the output. Dosovitskiy et al. (2015) designed a deep learning model that can be applied to generate photorealistic images of chairs from high-level descriptions (e.g., chair type and the viewpoint position). The model developed as a part of the present study mirrors the overall structure of the method proposed by Dosovitskiy et al. (2015). Specifically, high-level descriptions of browning apparatus conditions and dough sample serve as model inputs, generating an RGB image of browned dough as the output. To the best of the authors' knowledge, this is the first attempt to apply deep generative networks to food science. A further contribution of the present study stems from modeling images of browned products at a photorealistic, per-pixel RGB level, for the first time.

Physics-based CAD is increasingly being explored in food science research (Datta, 2016; Halder et al., 2011). The label-to-image browning prediction framework proposed in this work thus contributes to this effort by enabling further advancements in the CAD software aimed specifically at browning. Because the framework can yield photorealistic browned dough surfaces, it significantly improves CAD's ability to visualize final cooked dough samples prior to processing, while providing a more lifelike rendering for the user (Fig. 1). Such a visualization ability, while important for CAD applications, is impossible to achieve by applying previously discussed modeling methods that only characterize surface color with a single scalar value.

To test the robustness of the proposed label-to-image browning model, it was applied to a laser-induced browning pipeline. Laser-browning is a novel cooking technique that utilizes heat produced by a laser beam to induce browning in dough products (Blutinger et al., Under second round review). Similar to oven baking, as dough is heated by the laser beam, chemical reactions occur in the food product, giving rise to Maillard browning and caramelization (Blutinger et al., Under second round review). According to Martins et al. (2000), the non-enzymatic browning degree due to the Maillard reaction is related to aroma, taste, color, and nutritional value of cooked food products. Thus, it is important to precisely control the conditions that give rise to the final laser-cooked dough products, as this aids in achieving optimal browning patterns and eliminating any potentially harmful byproducts (Mottram et al., 2002).

Lasers differ from conventional heating methods in that they can provide high-resolution and controlled targeting of heat (Blutinger et al., 2018). These characteristics make them particularly appropriate for three-dimensional (3D) food printing application, which is a type of additive manufacturing (AM) technique that creates 3D edible objects by layering precisely defined two-dimensional (2D) slices of food material. Accurate targeting of heat is required in such applications due to the machine's ability to combine foods at a much closer level and the higher degree of customization that lasers offer. Additionally, the ability to precisely control the laser speed, power, flux, and cooking pattern renders this technique adaptable to different food materials (Blutinger et al., 2018). Due to the number of variables that need to be parameterized, laser browning is an ideal candidate for testing the proposed data-driven method's ability to model the browned dough characteristics.

2. Materials and methods

As illustrated in Fig. 3, the visual modeling pipeline is threefold: (1) experimental data must first be obtained (sections 2.1–2.3), (2) a deep generative network is then trained and tested using this data (sections 2.4–2.6), and (3) images generated by the trained and validated network are rendered to give a 3D photorealistic view (section 2.7).

2.1. Sample preparation

All-purpose flour (Gold Medal, General Mills, Minneapolis, USA) was used for the dough mixture. Five dough mixtures were prepared, each with a different flour to water ratio. The dough only consisted of flour and water. Moisture levels in the dough samples ranged from 50% to 70% water on a flour weight basis, in 5% increments (i.e., 50%, 55%, 60%, 65%, and 70%). Dough was mixed in a food processor (FP-8FR series, Cuisinart, East Windsor, USA) for 60 s at low speed under ambient conditions (23°C) and was left to rest for 15 min at 4°C. Dough samples were flattened into 2 mm (± 0.1 mm) thick rectangles prior to laser heating.

2.2. Laser apparatus

A CO_2 laser cutter and engraver (Nova 35, Thunder Laser Equipment Co., Ltd., Dongguan, China) was used to heat the dough. The laser operates at an infrared (IR) wavelength of 10.6 μm with a

Fig. 1. Rendered illustration of laser-browned crackers. The visual appearance of these crackers is fully predicted by the trained deep generative network.

maximum operating power of 80 W. RD Works V.8 software was used to adjust power, speed, and raster pattern. An "x-swing" scan mode (Fig. 4D) was employed in all experiments, with a 0.1 mm interval between neighboring scan passes. The software allowed for a maximum laser speed of 1000 mm s^{-1}, and a power range of W. Furthermore, according to the manufacturer data, the power consumption of the 80 W laser is 1400 W.

2.2.1. Browning the dough

Square raster cooking patterns (Fig. 4D) were laser-etched onto the dough while speed, laser energy flux, and dough moisture were adjusted for individual tests. Laser power was maintained at 8 W for all tests. Speed was controlled via the RD Works software, while energy flux was manually adjusted between tests by moving the laser bed (thus modifying the z-height). Since the Nova 35 laser cutter beam is not collimated, z-heights exceeding the focal length (4.5 mm) result in a lower laser energy flux, as the beam starts to diverge and increase in size. Each heated raster pattern spanned a 0.25 in × 0.5 in dough surface area. Based on the camera's pixel density, this area was sufficient to capture a browning pattern from the laser-heating process.

2.3. Data acquisition

Images were taken using a digital single-lens reflex camera (EOS Rebel T5i, Canon, Tokyo, Japan). Consistent lighting was ensured for all images. Placement of the light source was strategically placed at an oblique angle above the setup to minimize dough surface reflectance into the camera lens and accentuate surface patterns. Samples were imaged within minutes of laser browning the dough as to ensure surface quality and consistency. Additionally, all images were captured consecutively throughout the course of the day, so that the setup was not altered in any way. The captured images were post-processed and cropped to a 64 × 64 pixel size using MATLAB (Fig. 4A, B, & C). Experimentally obtained data with different dough moisture (w, %), laser height (z, mm), and laser speed (s, mm s^{-1}) is shown in Fig. 5.

2.4. Splitting training data and testing data

In order to demonstrate the network's robustness, the experimental data was separated into a training set and a testing set—not involved in training but is used for network validation—based on the following criteria:

- Dough moisture value (w): Data pertaining to w = 50%, 60%, and 70% was assigned to the training set (108 images), whereas that pertaining to w = 55% and 65% was used in testing (78 images). This separation permitted testing the network's ability to interpolate between the data points in the training dataset.
- Laser height (z): In this case, data with alternate z values was assigned to the training (96 images) and the testing datasets (90 images), respectively.
- Laser speed (s): Data with alternate s values was assigned to the

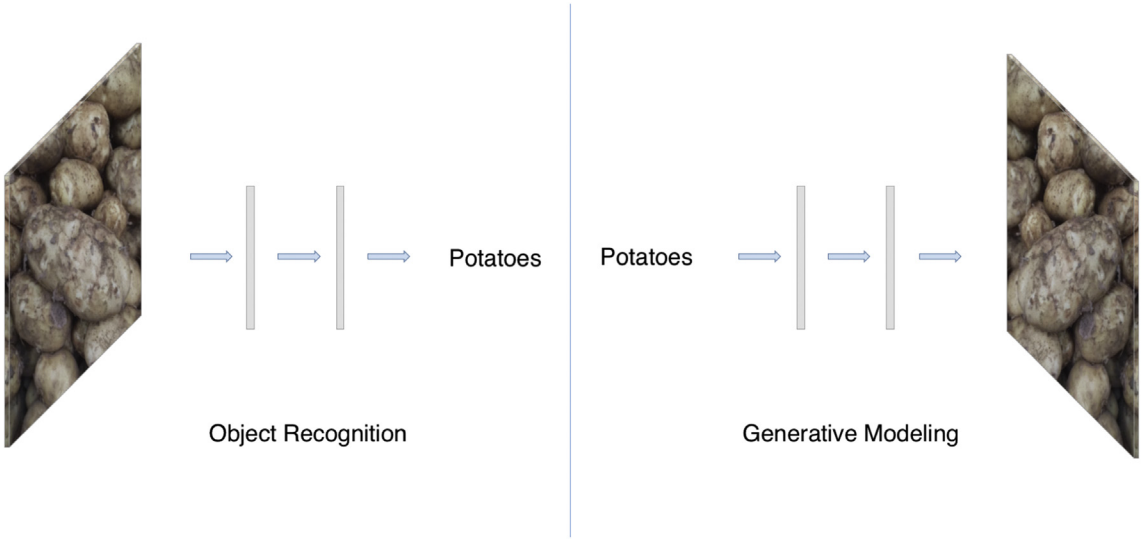


Fig. 2. Object recognition (label what is in the picture) vs. generative modeling (generate a picture from the label.).

Fig. 3. Visual modeling pipeline.

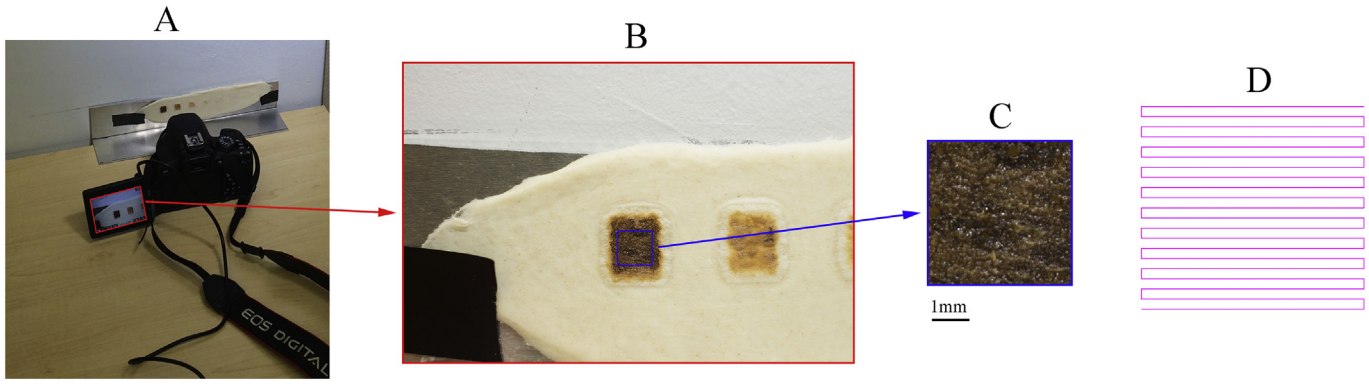


Fig. 4. Method of data acquisition. A) Camera setup to capture laser-browned dough. B) Sample image. C) 64 × 64 pixel size square was cropped from each image to be used for the network. D) Raster scan cooking pattern of the laser.

training (93 images) and the testing datasets (93 images), respectively.

- Checkerboard splitting: Instead of focusing on *w*, *z*, and *s* separately, alternate data points in the dataset as a whole were assigned to the training (93 images) and the testing datasets (93 images), respectively.

2.5. Data preprocessing for training

To use the hyperbolic tangent operator (tanh) as an activation function for the network, every RGB channel, which has a permissible intensity range from 0 to 255, of every pixel was normalized to the $[-1, +1]$ range. To enhance the numerical conditioning for the

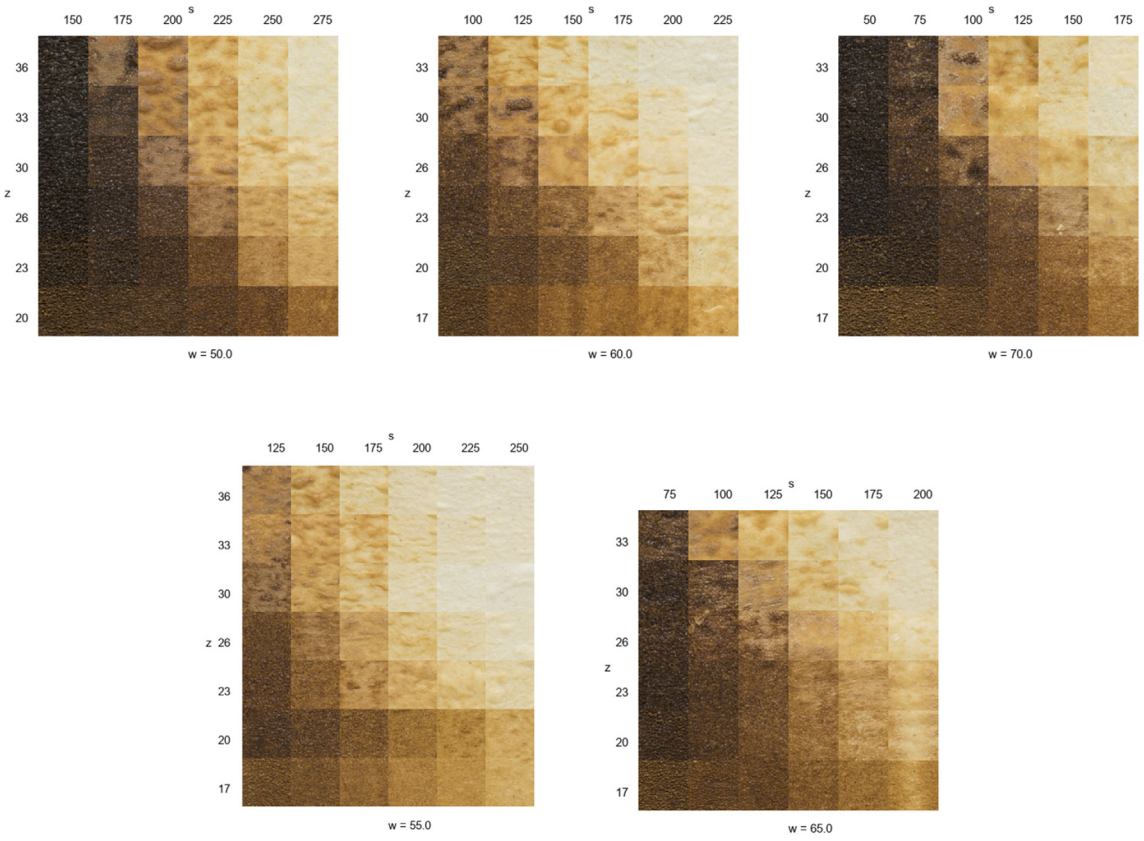


Fig. 5. Experimental Data. 186 unique images of laser-browned dough were obtained by varying dough moisture, laser height, and laser speed.

optimization solver during training, input variables (dough moisture, laser height, and laser speed) from the training dataset were standardized to have zero mean and unit standard deviation. No testing data was involved during this preprocessing. In the testing stage, we applied the same preprocessing algorithm to the testing dataset using mean and standard deviation of the training data.

2.6. Deep learning model description

The aim of the present investigation was to train a neural network that generates photorealistic RGB images of laser-browned dough from three scalar inputs, [w, z, s]. As discussed in Fig. 4 and section 2.4, w denotes the dough moisture value (water content), z represents laser height (directly affecting laser energy flux), and s indicates laser speed. RGB was chosen for the target image output format, since it matches the direct output format of the camera ($L^*a^*b^*$ would also be appropriate for characterizing the ground truth images). For simplicity, henceforth, three scalar numbers listed in square brackets indicate a set of w, z, and s values. For example, [55, 26, 150], denotes w = 55%, z = 26 mm, s = 150 mms^{-1}.

2.6.1. Data augmentation

We applied data augmentation to the training dataset to reduce the likelihood of overfitting during training and increase the network's ability to generalize (Wang and Perez, 2017).

Specifically, the training data was augmented by introducing two random scalars θ_x, artificial horizontal translation, and θ_y, artificial vertical translation (Fig. 6). These two randomized scalars were concatenated to the original three input variables, [w, z, s], to form a new input vector, [w, z, s, θ_x, θ_y].

Each training image was augmented via 20 artificial random translations, effectively increasing the size of the training dataset by a factor of 20. Thus, the training dataset grew from 108 images to 2160 images in the case when training and testing data are separated based on dough moisture value (w); from 96 images to 1920 images when data was separated based on laser height (z); from 93 images to 1860 images when data was separated based on laser speed (s); and from 93 images to 1860 images when data was split based on the checkerboard pattern.

2.6.2. Network architecture

Convolutional neural networks (CNNs), a major reason behind the robust performance of object recognition problems, are optimized for processing data presented in the form of multiple arrays, such as RGB images (LeCun et al., 2015). In contrast to object recognition, deconvolution networks—used for generative modeling—can produce RGB images from label parameters (w, z, and s) (Dumoulin and Visin, 2016; Zeiler et al., 2010). A deconvolution network architecture (Fig. 7a) similar to that proposed by Dosovitskiy et al. (2015) was adopted in the present study to generate photorealistic laser-browned dough images.

The input to the network are physical parameters [w, z, s] and the horizontal and vertical data augmentation parameters [θ_x, θ_y]. Through several fully connected layers and multiple deconvolution layers, the network gradually increases the dimension of the output until finally a 64 × 64 image of three channels (RGB image) was obtained. The generated image was then compared with the ground truth image to evaluate the loss function during training. The loss function measures the average difference per RGB channel per pixel between the generated image and the ground truth image.

Although, in this particular setup, the output image resolution was 64 × 64, deconvolution layers can be added or removed to generate a higher or lower image resolution, depending on the application and available computational resources. For example, Wang et al. (2017) recently used deep generative network to generate images of 2048 × 1024 resolution.

In addition to the above structure, a batch normalization layer was added after each layer. Batch normalization layer improves the training process by optimally normalizing the output of its previous layer (Ioffe and Szegedy, 2015). Even though batch normalization was not employed by Dosovitskiy et al. (2015) in their work, it is prevalent in deep generative network studies (Radford et al., 2015). Additional training details are listed in Appendix B.

2.7. Rendering tools

Quality visualization techniques are crucial for CAD applications since they allow for a more detailed and accurate communication of information. After training, the deep generative network is able to predict 2D images of browned dough. Blender, an open-source 3D computer graphics software (Blender, 2018), was then used to post-process these images by

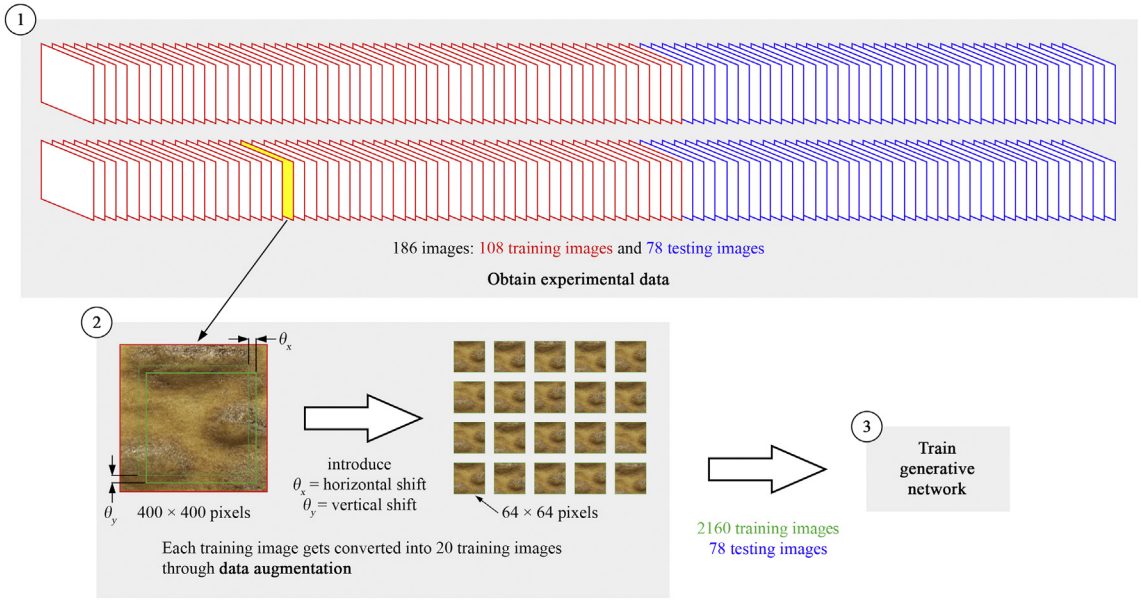
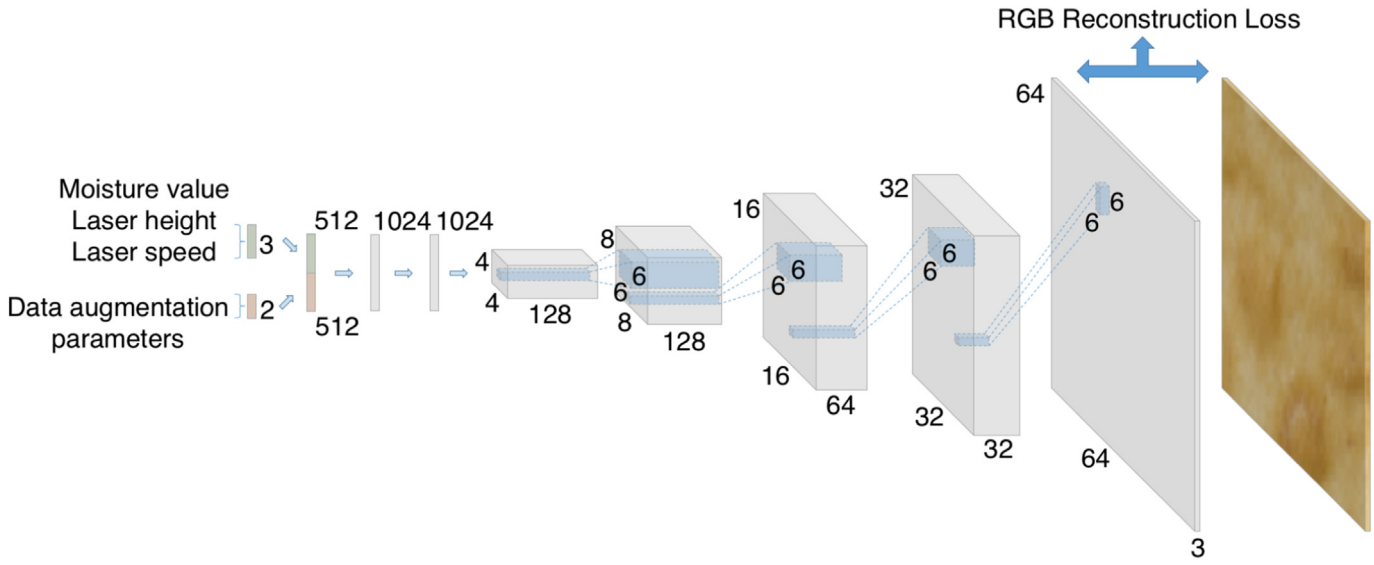
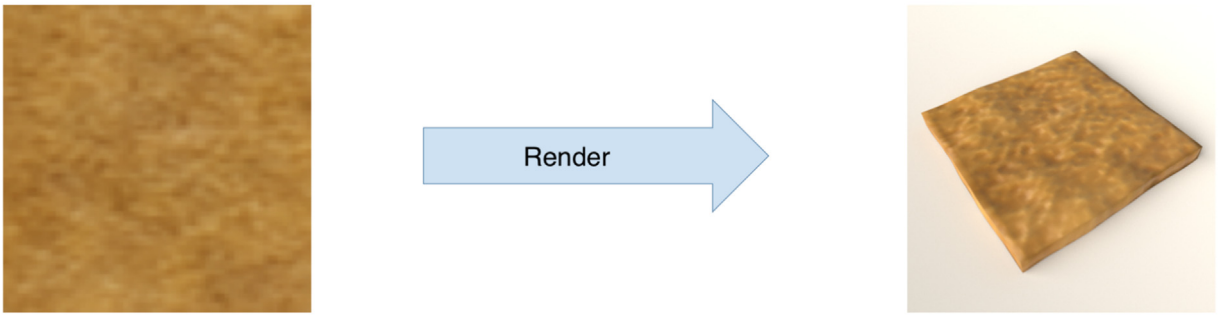


Fig. 6. Data augmentation method. (1) 186 images were captured, 58% for training and 42% for testing. (2) Each training image was augmented 20 times using artificial translations. (3) After data augmentation, 2160 images were available for training and 78 images for testing.



(a) Architecture that generates a 64 x 64 pixel image from laser and dough parameter inputs.



(b) A 2D prediction image generated by the network is rendered to give a 3D illustration.

Fig. 7. Network architecture (a) and rendering pipeline (b).

texture-mapping them onto the surface a 3D object (Figs. 7b and 8d).

3. Results and discussion

The network's ability to generate predictive images of laser-browned dough after training on different training datasets was evaluated. Batch normalization was incorporated into all network layers, unless otherwise noted. In this section, only the results obtained by separating the training and testing data based on dough moisture value (w) are shown, while those obtained when data was separated based on laser height (z), laser speed (s), and using the checkerboard method, are included in the appendix.

To demonstrate the network's ability to learn dough moisture (w), as previously indicated in section 2.4, the network was trained using data related to dough with w = 50%, 60%, and 70% (Fig. 8a), and was tested using images related to w = 55% and 65% (Fig. 8b). It should be emphasized that none of the testing data (ground truth) images were available to the network during training (section 2.4).

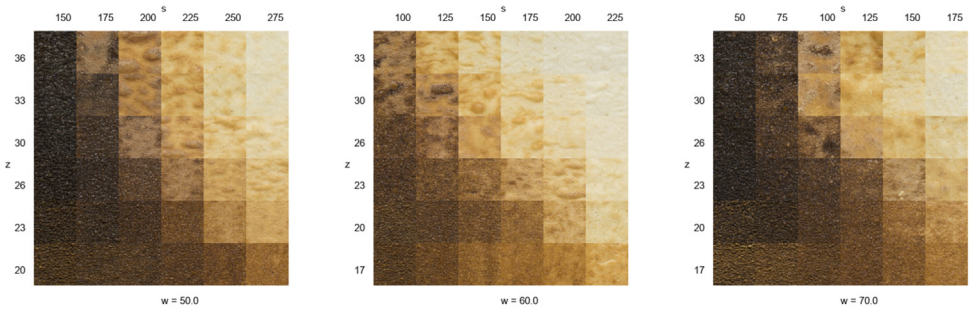
Qualitatively, the prediction images generated by the network after training (Fig. 8c) are not only photorealistic, but also match ground truth (Fig. 8b) images for both w = 55% and w = 65%. Both ground truth and generated images exhibited a darker color and high-frequency dotted patterns when low z and s values were used. As z and s increased, resulting in a lower laser energy flux and lower exposure time, surface color became gradually lighter, with surface patterns of lower frequencies. Fig. 8d gives a rendered view of the lower right corner nine images of w = 65% in Fig. 8c, i.e. z = 17, 20, 23 mm, s = 150, 175, 200 mm s^{-1}, hence demonstrating the ability to use images generated by the network to develop 3D rendered visuals (rendered in Blender).

Quantitatively, the average error per RGB channel per pixel of the generated image was 8.02% of the permissible intensity range (0 − 255).

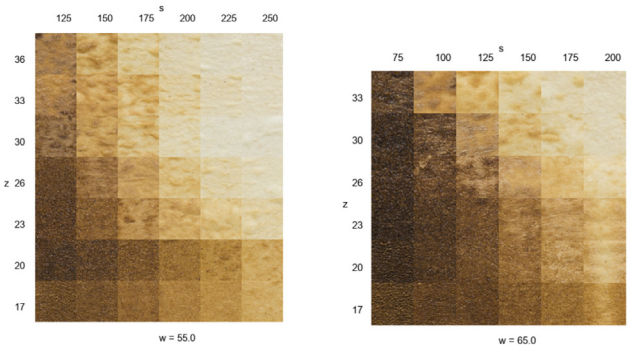
Even though the generated images predict a correct general trend, upon close inspection of individual images, it was evident that the visual details in most images predicted by the network differed from the corresponding ground truth. The generated images also lacked detailed features and were thus visually not as “sharp” as their ground truth equivalents.

It should also be noted that quality of the generated images can vary greatly between contiguous image tiles. For example, when [55, 30, 125] (i.e., w = 55%, z = 30 mm, and s = 125 mm s^{-1}) is chosen, the generated image and the ground truth share similar dark black color and pattern. However, when only s is changed to 150, i.e., when [55, 30, 150] is used, the generated image and the ground truth image have clearly different color.

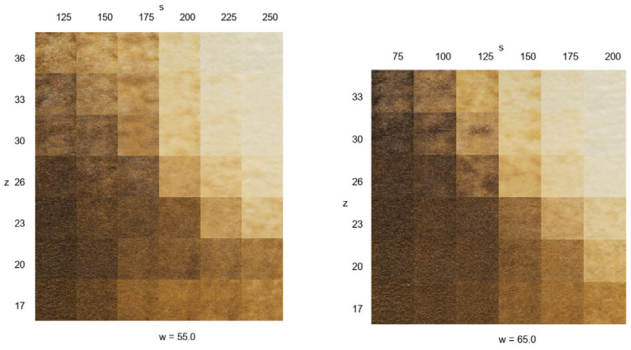
This label-dependent consistency issue likely arises because not enough sampling was applied to the training data. In the training dataset (Fig. 8a), the image tile for [55, 30, 150] has a very dark color and consistent surface irregularity (high frequency pattern), while the image tile for [60,30,150] has a white/yellow color and more significant textural features such as air bubbles (low frequency pattern). Such drastic differences (discontinuities) in image features between these two adjacent training images pose a serious challenge for the generative network's capacity to predict any intermediate values, such



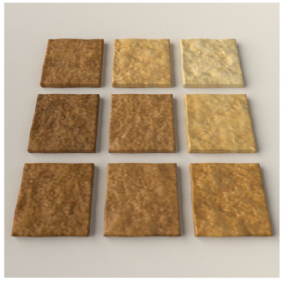
(a) Training data



(b) Testing data



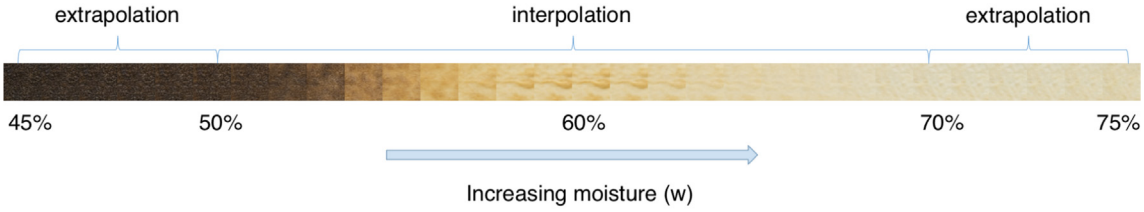
(c) Prediction generated by the network



(d) Rendered view of the lower right corner nine images of $w = 65\%$ in (c)

Fig. 8. Splitting training and testing data based on dough moisture value (w).

(a) Training data, $w = 50\%$ (b) Training data, $w = 60\%$ (c) Training data, $w = 70\%$



(d) Images generated with deep learning from $w = 45\%$ to $w = 75\%$ with 1.03% difference between neighboring images.

Fig. 9. Training data and generated images of z = 30 mm and s = 175 mm s^{-1}.

as [55, 30, 150]. Thus, further training data is required to address the discontinuities among adjacent training images and ensure more accurate interpolation output. Adaptive sampling techniques for unbalanced datasets could also be incorporated to efficiently mitigate this problem (He et al., 2008).

3.1. Continuous interpolation between images

While the interpolation results discussed in the preceding section pertained to the interpolation at ground truth label, the focus on this section is on interpolation at arbitrary input value, thereby achieving interpolation in a continuous range. Continuous interpolation is a common feature needed by CAD applications because users often want to try different input values and see what the manufactured product virtually looks like before finalizing the input parameters.

Fig. 9a, b, & c shows images of training data of laser height z = 30 mm and laser speed s = 175 mm s^{-1}. Only three dough moisture values (w) associated with this z and s pair were present in the training dataset (50%, 60%, 70%). Nonetheless, after training, the deep learning model was able to generate an arbitrary combination of [w, z, s]. Fig. 9d shows the predicted images for z = 30 mm and laser speed s = 175 mm s^{-1} with varying dough moisture values. In the interpolation region, the dough moisture values span from 50% to 70% water on flour weight basis, with 1.03% difference between neighboring images. A continuous transition in color (from black to brown to white), and in surface pattern (from more uniform surface pattern, to less consistent surface bumps, to smooth) can be observed. This result demonstrates that the deep neural network did not merely memorize the training data (overfit), but fitted it in a generalized manner that enables continuous nonlinear interpolation between the data points in the training set. The network can also perfectly reproduce the training images pertaining to the 50%, 60%, and 70% dough moisture content (Fig. 9d).

3.2. Assessing network's potential for extrapolation

To test the network's robustness, dough moisture values (w) outside of the training data's range, specifically its convex hull (50 − 70%), were input into the trained model in order to assess the network's ability to extrapolate images from untested sample sets (Fig. 10a and c). Similar to the settings used in the tests discussed in the preceding section, both laser height (z) and speed (s) were held constant at 30 mm and 175 mm s^{-1}, respectively.

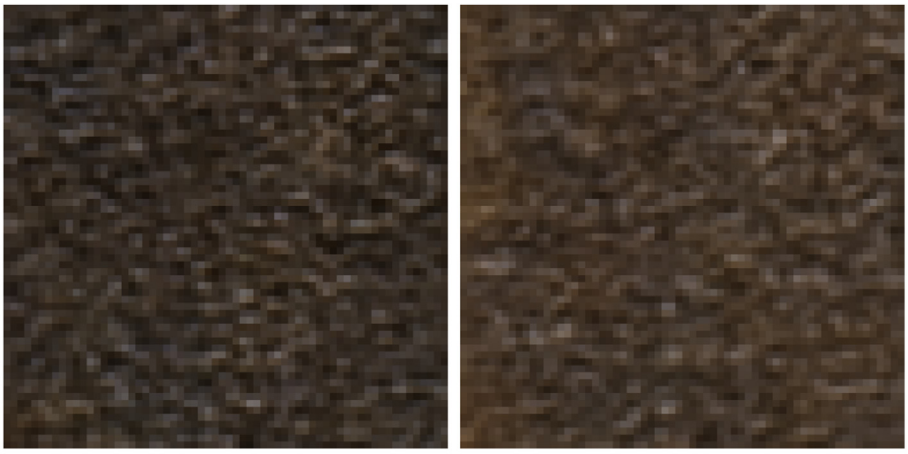
In these experiments, images for less saturated dough (w = 45%) and more saturated dough (w = 75%) were generated to assess the network's ability to extrapolate (Fig. 10a and c) while surface images of laser-browned dough with w = 50% and w = 70% were used for comparison (Fig. 10b and d). The findings indicated that, despite the absence of ground truth images for w = 45% and w = 75%, the generated images were photorealistic and similar to their nearest-neighbor images in the training dataset. Such results confirm the deep generative network's capability to extrapolate photorealistic images.

Continuous extrapolation was also experimented and demonstrated in the extrapolation regions of Fig. 9d. Dough moisture values span from 45% to 50% water on flour weight basis and then from 70% to 75% water on flour weight basis with 1.03% difference between neighboring images.

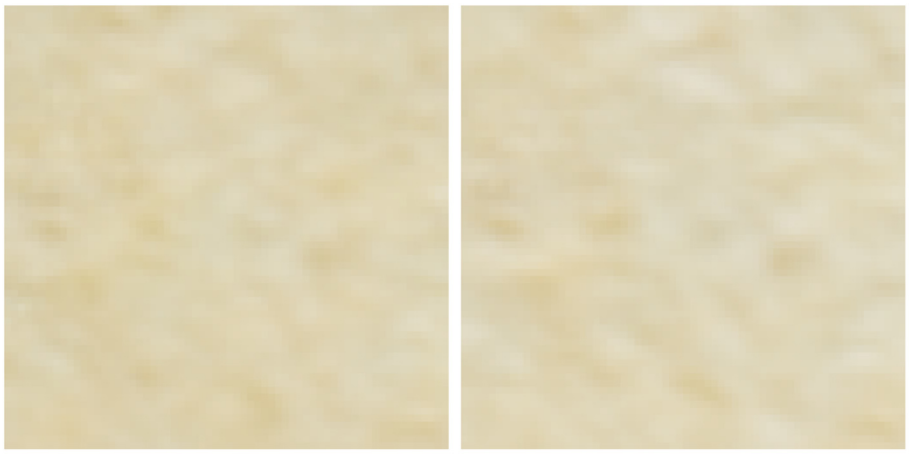
While the network can produce reasonable images through extrapolation, its ability to extrapolate realistic images outside of the established “cookable” range of dough requires further ground-truthing. From initial benchmark testing, if dryer dough (low w) is heated at a low laser speed, and high laser energy flux (low z), the surface of the exposed dough gets vaporized, transitions past burning (dark color), and produces a powdery unappetizing end result (lighter color). Had this cooking phenomenon been well documented in the training set, than the network would have been able to extrapolate more realistic results.

3.3. Effect of batch normalization

Fig. 11 shows a comparison between the images generated with and without batch normalization (BN), indicating that the latter suffer from an artifact due to processing that resembles white noise, especially at

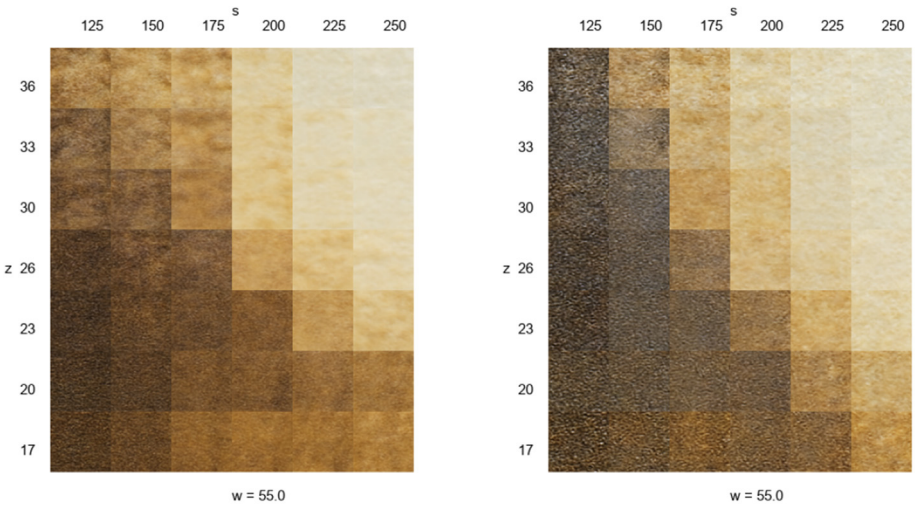


(a) $w = 45\%$ (outside training data range) (b) $w = 50\%$ (inside training data range)



(c) $w = 75\%$ (outside training data range) (d) $w = 70\%$ (inside training data range)

Fig. 10. Generated images at extreme dough moisture values (w).



(a) with BN (b) without BN

Fig. 11. Generated images with and without BN.

lower laser speeds. Average error per RGB channel per pixel of all the generated images without BN is 8.25% of the permissible intensity range $(0 - 255)$, which declines to 8.02% when BN is employed. Even though this quantitative improvement seems negligible, the visual differences are apparent. This mismatch between visual perception and quantitative error indicates that the error function needs to be refined so that it better reflects the visual perception.

Throughout the present study, a Euclidean norm (average per RGB channel per pixel loss function) was employed to measure the difference between the generated image and the ground truth image (c.f., loss function in section 2.6.2 and Appendix B). Advantages to using a Euclidean norm is that it is straightforward to implement and has a stable training process. While this approach can lead to a trained model that generates photorealistic images, it does not penalize for digital white noise artifacts (visible in Fig. 11b). Future research could explore alternative measurement schemes such as SSIM (Wang et al., 2004) to compare clusters of pixels rather than single pixel values. Furthermore, recent work on the generative adversarial network (GAN) demonstrates that promising results can be obtained by training another deep neural network to measure feature similarity between images (Goodfellow et al., 2014; Radford et al., 2015). The GAN approach can also be beneficial for mitigating the previously discussed sharpness issue (Isola et al., 2017).

4. Conclusions

In this work, we propose a visual modeling pipeline of dough browning. Experimentally obtained images of laser-browned dough were used to train a deep generative network. After training, our network was able to predictively generate photorealistic images of browned dough, which can be textured onto a 3D model. This exploration—being the first of its kind—will serve as a benchmark for further visual modeling studies of food and will allow designers to produce more accurate and photorealistic food products.

The architecture of our data-driven supervised learning model also makes it easily extensible to other labeled image datasets (i.e., traditional oven baking, frying, drying, etc.). Future studies should investigate alternative loss functions and training data sampling schemes as a means of furthering these results.

Acknowledgements

This work was supported in part by Columbia University's SEAS Interdisciplinary Research Seed (SIRS) funding program.

Appendix A. Results for splitting training and testing data based on laser height (z), laser speed (s), and checkerboard

Appendix A.1. Effect of splitting data based on laser height (z)

As shown in Figure A.1, average error per RGB channel per pixel of all the generated images is 7.38% of the permissible intensity range $(0 - 255)$.

Appendix A.2. Effect of splitting data based on laser speed (s)

As shown in Figure A.2, average error per RGB channel per pixel of all the generated images is 7.06% of the permissible intensity range $(0 - 255)$.

Appendix A.3. Effect of checkerboard-splitting of data

As shown in Figure A.3, average error per RGB channel per pixel of all the generated images is 7.16% of the permissible intensity range $(0 - 255)$.

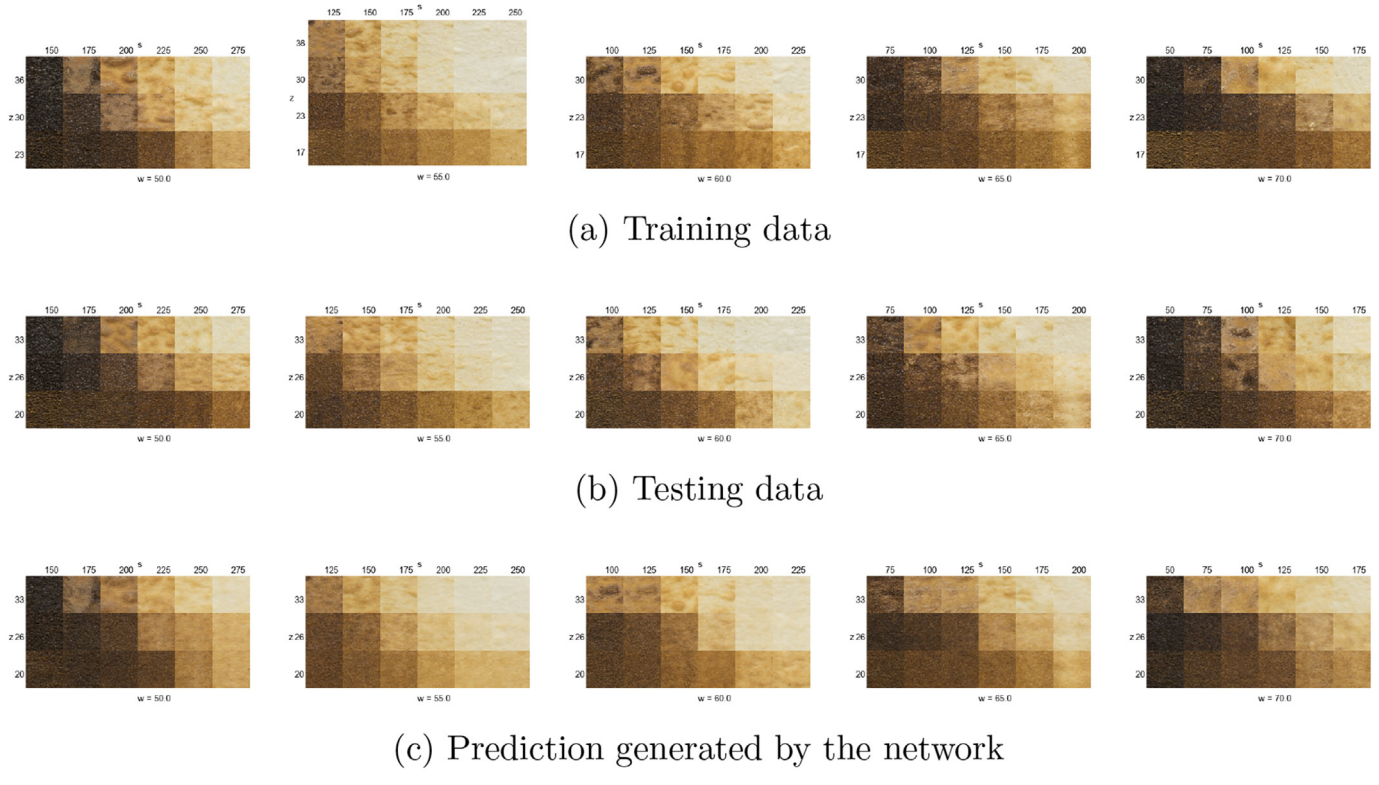
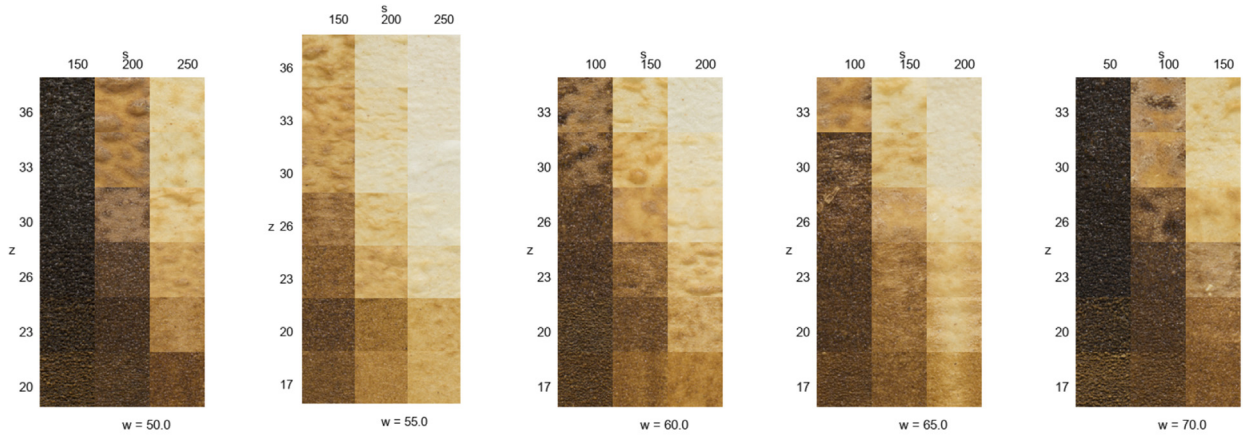
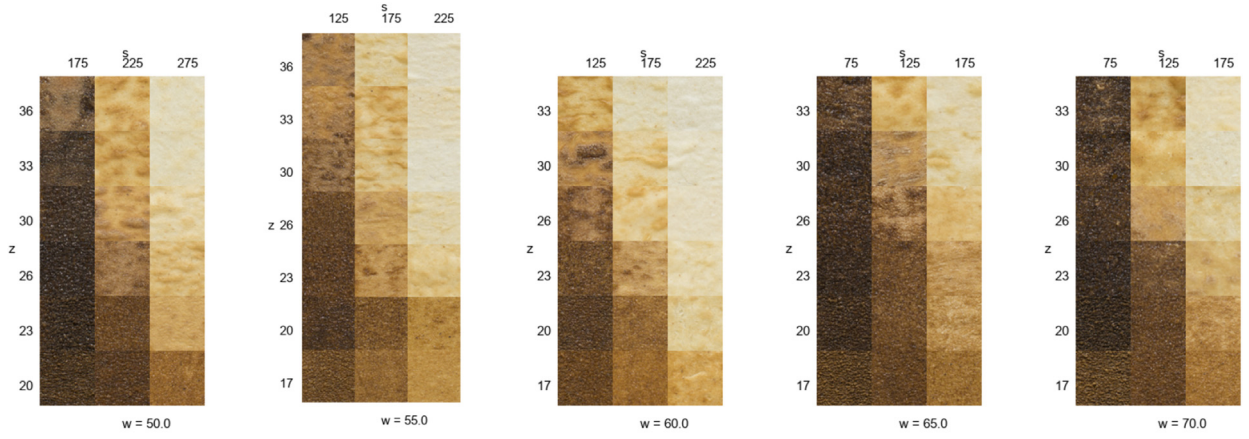


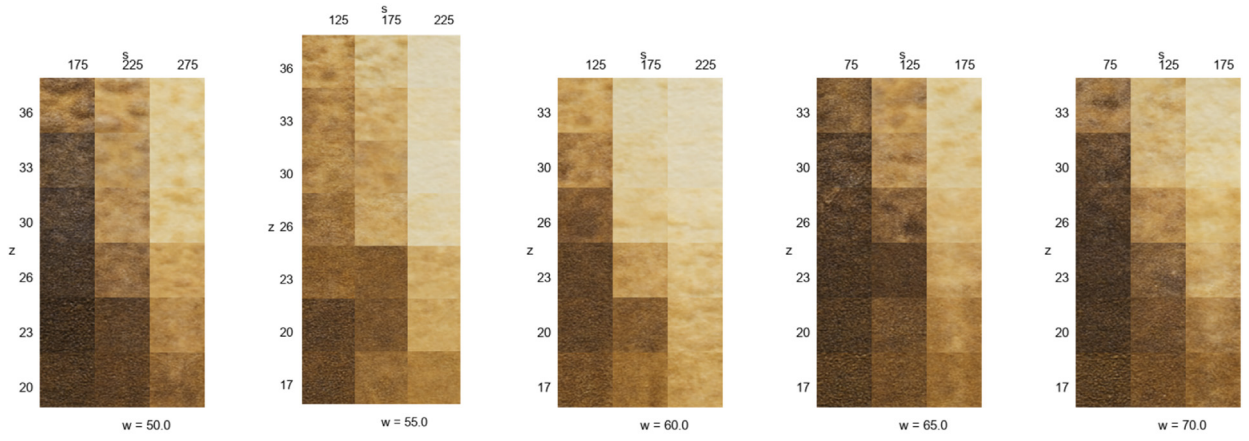
Fig. A.1. Splitting training and testing data based on laser height (z).



(a) Training data



(b) Testing data



(c) Prediction generated by the network

Fig. A.2. Splitting training and testing data based on laser speed (*s*).

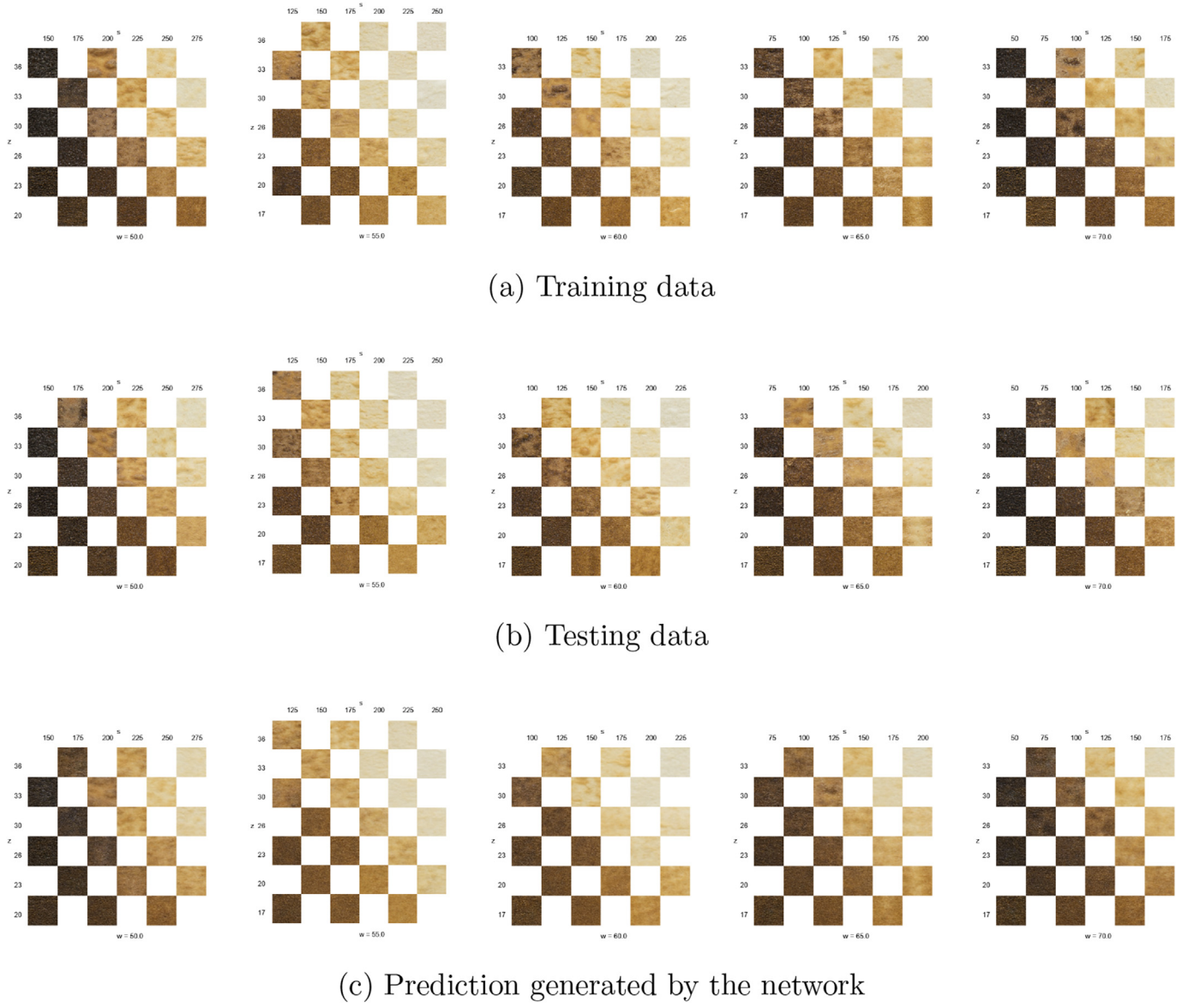


Fig. A.3. Checkerboard-splitting training and testing data based.

Appendix B. Training details

The network architecture described in section 2.6.2 was implemented in Keras (Chollet et al., 2015) with the following details:

- Deconvolution layer parameters: kernel size (6) and stride size (2) were fixed across the network. These parameters can also be perturbed, but no significant change to the quality of the network should be observed.
- Activation function: ReLU (Nair and Hinton, 2010) was used as the activation function for all but the last layer, where tanh was used.
- Weight initialization: The normal initiator (He et al., 2015) was used for initializing network weights of all layers.
- Loss function: L_1 Euclidean norm or mean absolute error (MAE) was used to measure the difference between the generated and the ground truth image per RGB channel per pixel. Although L_2 Euclidean norm or mean squared error can also be used, no significant changes in the network quality should be observed.
- Optimization solver: Adam (Kingma and Ba, 2014) with momentum parameters $\beta_1 = 0.9$, $\beta_2 = 0.999$ and regularization parameter $\varepsilon = 10^{-8}$ was used. A constant learning rate of 0.0002, constant batch size of 32, and epoch size of 10000 were chosen for all the examples discussed in the following sections.

References

Blender, 2018. Blender - a 3d modelling and rendering package. http://www.blender.org.
Blutinger, J.D., Meijers, Y., Chen, P.Y., Zheng, C., Grinspun, E., Lipson, H., 2018. Characterization of dough baked via blue laser. J. Food Eng. 232, 56–64. http://www.sciencedirect.com/science/article/pii/S026087741830133X.
Blutinger, J. D., Meijers, Y., Chen, P. Y., Zheng, C., Grinspun, E., Lipson, H., Under Second Round Review. Characterization of Co2 Laser Browning of Dough. Innovat. Food Sci. Emerg. Technol.
Chollet, F., et al., 2015. Keras: deep learning library for theano and tensorflow. https://keras.io/k7,8.

Datta, A., 2016. Toward computer-aided food engineering: mechanistic frameworks for evolution of product, quality and safety during processing. J. Food Eng. 176, 9–27.
DeChant, C., Wiesner-Hanks, T., Chen, S., Stewart, E.L., Yosinski, J., Gore, M.A., Nelson, R.J., Lipson, H., 2017. Automated identification of northern leaf blight-infected maize plants from field imagery using deep learning. Phytopathology 107 (11), 1426–1432.
Dosovitskiy, A., Tobias Springenberg, J., Brox, T., 2015. Learning to generate chairs with convolutional neural networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1538–1546.
Du, C.-J., Sun, D.-W., 2006. Learning techniques used in computer vision for food quality evaluation: a review. J. Food Eng. 72 (1), 39–55.
Dumoulin, V., Visin, F., 2016. A Guide to Convolution Arithmetic for Deep Learning. arXiv preprint arXiv:1603.07285.
Goodfellow, I., Bengio, Y., Courville, A., 2016. Deep Learning. MIT press.
Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y., 2014. Generative adversarial nets. In: Advances in Neural Information Processing Systems, pp. 2672–2680.
Halder, A., Dhall, A., Datta, A.K., Black, D.G., Davidson, P., Li, J., Zivanovic, S., 2011. A user-friendly general-purpose predictive software package for food safety. J. Food Eng. 104 (2), 173–185.
He, H., Bai, Y., Garcia, E.A., Li, S., 2008. Adasyn: adaptive synthetic sampling approach for imbalanced learning. In: Neural Networks, 2008. IJCNN 2008.(IEEE World Congress on Computational Intelligence). IEEE International Joint Conference on. IEEE, pp. 1322–1328.
He, K., Zhang, X., Ren, S., Sun, J., 2015. Delving deep into rectifiers: surpassing human-level performance on imagenet classification. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 1026–1034.
Ioffe, S., Szegedy, C., 2015. Batch normalization: accelerating deep network training by reducing internal covariate shift. In: International Conference on Machine Learning, pp. 448–456.
Isola, P., Zhu, J.-Y., Zhou, T., Efros, A.A., 2017. Image-to-image Translation with Conditional Adversarial Networks. arXiv preprint.
Kingma, D.P., Ba, J., 2014. Adam: a Method for Stochastic Optimization. arXiv preprint arXiv:1412.6980.
LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521 (7553), 436–444.
Martins, S.I., Jongen, W.M., Van Boekel, M.A., 2000. A review of maillard reaction in food and implications to kinetic modelling. Trends Food Sci. Technol. 11 (9–10), 364–373.
Mezgec, S., Koroušić Seljak, B., 2017. Nutrinet: a deep learning food and drink image recognition system for dietary assessment. Nutrients 9 (7), 657.
Mottram, D.S., Wedzicha, B.L., Dodson, A.T., 2002. Food chemistry: acrylamide is formed in the maillard reaction. Nature 419 (6906), 448.
Nair, V., Hinton, G.E., 2010. Rectified linear units improve restricted Boltzmann machines. In: Proceedings of the 27th International Conference on Machine Learning (ICML-10), pp. 807–814.
Pfisterer, K.J., Amelard, R., Chung, A.G., Wong, A., 2018. A new take on measuring relative nutritional density: the feasibility of using a deep neural network to assess commercially-prepared puréed food concentrations. J. Food Eng. 223, 220–235 Elsevier.
Purlis, E., Salvadori, V.O., 2007. Bread browning kinetics during baking. J. Food Eng. 80 (4), 1107–1115.
Purlis, E., Salvadori, V.O., 2009. Modelling the browning of bread during baking. Food Res. Int. 42 (7), 865–870.
Radford, A., Metz, L., Chintala, S., 2015. Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks. arXiv preprint arXiv:1511.06434.
Sablani, S.S., 2008. Status of observational models used in design and control of products and processes. Compr. Rev. Food Sci. Food Saf. 7 (1), 130–136.
Wang, J., Perez, L., 2017. The effectiveness of data augmentation in image classification using deep learning. Tech. rep (Technical report), arXiv preprint arXiv:1712.04621.
Wang, T.-C., Liu, M.-Y., Zhu, J.-Y., Tao, A., Kautz, J., Catanzaro, B., 2017. High-resolution Image Synthesis and Semantic Manipulation with Conditional Gans. arXiv preprint arXiv:1711.11585.
Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P., 2004. Image quality assessment: from error visibility to structural similarity. IEEE Trans. Image Process. 13 (4), 600–612.
Zanoni, B., Peri, C., Bruno, D., 1995. Modelling of browning kinetics of bread crust during baking. LWT - Food Sci. Technol. 28 (6), 604–609.
Zeiler, M.D., Krishnan, D., Taylor, G.W., Fergus, R., 2010. Deconvolutional networks. In: Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on. IEEE, pp. 2528–2535.
Zhang, J., Datta, A., 2006. Mathematical modeling of bread baking process. J. Food Eng. 75 (1), 78–89.