Channel Randomisation with Domain Control for Effective Representation Learning of Visual Anomalies in Strawberries

Taeyeong Choi and Grzegorz Cielniak

Lincoln Agri-Robotics (LAR) Centre, University of Lincoln, UK
{tchoi, gcielniak}@lincoln.ac.uk

Abstract

Channel Randomisation (CH-Rand) has appeared as a key data augmentation technique for anomaly detection on fruit images because neural networks can learn useful representations of *colour* irregularity whilst classifying the samples from the augmented "domain". Our previous study has revealed its success with significantly more reliable performance than other state-of-the-art methods, largely specialised for identifying structural implausibility on non-agricultural objects (e.g., screws). In this paper, we further enhance CH-Rand with additional *guidance* to generate more informative data for representation learning of anomalies in fruits as most of its fundamental designs are still maintained. To be specific, we first control the "colour space" on which CH-Rand is executed to investigate whether a particular model — e.g., HSV, $YCbCr$, or $L^*a^*b^*$ — can better help synthesise realistic anomalies than the RGB, suggested in the original design. In addition, we develop a learning "curriculum" in which CH-Rand *shifts* its augmented domain to gradually increase the difficulty of the examples for neural networks to classify. To the best of our best knowledge, we are the first to connect the concept of curriculum to self-supervised representation learning for anomaly detection. Lastly, we perform evaluations with the Riseholme-2021 dataset, which contains $>$ 3.5K real strawberry images at various growth levels along with anomalous examples. Our experimental results show that the trained models with the proposed strategies can achieve over 16% higher scores of AUC-PR with more than three times less variability than the naïve CH-Rand whilst using the same deep networks and data.

Introduction

Reliable perception systems are essential to fully automate various tasks in agricultural applications. For instance, fruit monitoring robots must be able to not only sense individual instances of fruit but also precisely assess their quality for predicting future yield or performing targeted treatment depending on their health conditions. Collecting visual examples of anomalous cases — e.g., fruits with disease or damage — is, however, a challenging process in training deep networks generally because of their rare occurrences. Thus, One-class Classification (OC) can be a practical solution for learning to classify anomalies when only normal data is available during training (Choi et al. 2021b; Li et al. 2021; Choi et al. 2021a).

Figure 1: Examples from the Riseholme-2021 dataset. On the left three columns are image samples of normal strawberries at ripe and unripe stages with the possibility of occlusion, while the right two display anomalous examples.

Self-supervised Learning (SL) has been widely used to build high-performance anomaly detectors in OC, for which normal data are augmented in particular ways so that a deep network can gain informative representations for anomaly identification whilst learning to solve some *pretext* tasks with the augmented samples — e.g., rotation prediction (Hendrycks et al. 2019) and position inference with patches (Yi and Yoon 2020). For anomaly detection in *fruits*, Choi et al. (2021b) have introduced Channel Randomisation (CH-Rand) to augment images of healthy instances by randomly permuting values across the channel dimension. By classifying between the original and randomised fruit images, their model could learn representations of implausible "colour" patterns to outperform other state-of-the-art methods such as CutPaste (Li et al. 2021), designed to learn "structural" defects in non-agricultural items (e.g., screws, wires, and carpets in MVTec AD (Bergmann et al. 2021))

In this paper, we enhance the utility of CH-Rand with strategies to produce more useful data in augmentation for effective SL of anomalies in fruits. First of all, we perform *colour space* conversion with a hypothesis that CH-Rand on a particular colour model, such as HSV, $YCbCr$, or $L^*a^*b^*$, may better simulate the data *domain* of real anomalies than on the RGB space, explored earlier in (Choi et al. 2021b). In addition, inspired by (Bengio et al. 2009), we also build a learning "curriculum" to incrementally increase the difficulty of samples in the pretext task by regulating the set

of random channel sequences to consider. To the best of our knowledge, we are the first to utilise the concept of curriculum to improve SL for anomaly detection. Moreover, similar to (Choi et al. 2021b), all our experiments are set up to solve OC with *Riseholme-2021*, a large dataset of > 3.5K strawberry images, because it presents a realistically challenging testbed with examples at various maturity stages along with anomalous ones in wild conditions featuring frequent occlusions and varying illumination (cf. Fig. 1).

Methodology

In this section, we briefly describe the design of CH-Rand, proposed in (Choi et al. 2021b), and also introduce novel settings to alter the domains that CH-Rand can create to potentially benefit SL of visual anomalies in real strawberries.

Channel Randomisation

CH-Rand is an image augmentation method to encourage neural networks to learn to discern *unnatural* colour compositions from the normal ones. To achieve this goal, the augmentation is performed by randomly permuting values in the channel dimension with the possibility of repetition.

More formally, CH-Rand is set to map each normal image $\mathcal{I} \in \mathbb{R}^{W \times H \times C}$ to an augmented image $\mathcal{A}^{W \times H \times C}$, where W, H, and C are the dimensions of width, height, and colour channel, respectively. This transformation basically uses an arbitrary function $\pi : \chi \rightarrow \chi'$ for permutation, where $\chi = \{1, 2, ..., C\}$, and $\chi' \in \mathcal{P}(\chi) \setminus \varnothing$ with $\mathcal{P}(\cdot)$ as the powerset of input. Note that we avoid obtaining $\mathcal{A} = \mathcal{I}$ by drawing π satisfying $\exists c \in \chi, c \neq \pi(c)$. Hence, if $C = 3$, 26 distinct images can be constructed by channel sequences $(\pi(1), \pi(2), \pi(3))$ depending on π.

Eventually, the same function π is applied to compute every element $a^c_{w,h} \in \mathcal{A}$ from $\mathcal{I}$:

$$a^c_{w,h} = i^{\pi(c)}_{w,h}. \tag{1}$$

Similar to (Li et al. 2021; Gidaris, Singh, and Komodakis 2018), a classifier f_Θ can then be trained to classify the augmented images from the ones without augmentation minimising the loss function below:

$$\mathcal{L} = \mathbb{E}_{\mathcal{I} \in \mathcal{D}} \Big[H(f_\Theta(\mathcal{I}), 0) + H(f_\Theta(CHR(\mathcal{I})), 1) \Big], \tag{2}$$

where $\mathcal{D}$ denotes the set of normal images available for training in OC scenarios, and CHR and H are the application of CH-Rand and the binary cross entropy, respectively.

Anomaly Score Calculation

Choi et al. (2021b) have suggested utilising the feature space g_θ, which is generated by an intermediate layer within f_Θ, to calculate the anomaly score s for test input $\mathcal{I}'$. As in (Perera and Patel 2019), the mean *distance* from the k nearest neighbors $\mathcal{N}$ in training set $\mathcal{D}$ is considered — i.e. $s(\mathcal{I}') = (1/k) \sum_{\mathcal{I} \in \mathcal{N}} \delta\big(g_\theta(\mathcal{I}), g_\theta(\mathcal{I}')\big)$, where δ computes the *Euclidean* distance between two inputs — to determine the novel input $\mathcal{I}'$ as anomaly if s is larger than a particular threshold γ .

In this work, we keep all these fundamental schemes to only focus on the effect of altered data domains by the techniques discussed in the following section.

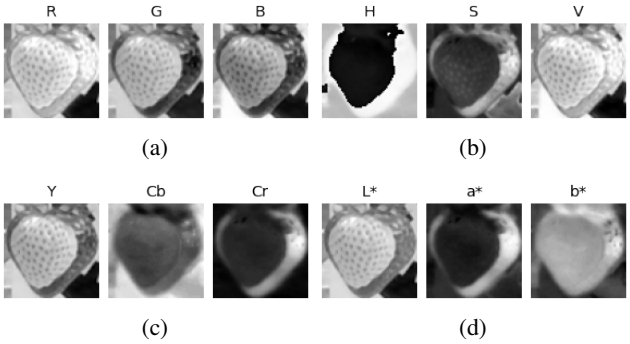


Figure 2: Individual channels in different colour spaces to express the image of normal strawberries in Fig. 3a: (a) RGB, (b) HSV, (c) $YCbCr$, and (d) $L^*a^*b^*$.

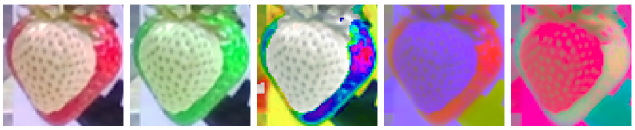


Figure 3: Examples of CH-Rand applied to a normal strawberry image in (a). (b)–(e) are the results of exchanging values between the first and the second channel in RGB, HSV, $YCbCr$, and $L^*a^*b^*$, respectively. Each outcome has been converted to the RGB after exchange for visualisation.

Domain Control Strategies

We here present two additional modules — 1) colour space conversion and 2) curriculum learning — for data augmentation to further improve anomaly detection on strawberry data. With these add-ons, the resulting images are expected to compose data domains where neural networks can better learn useful representations of anomalous strawberries.

Colour Space Conversion We hypothesise that CH-Rand on a particular colour space — e.g., HSV, $YCbCr$, or $L^*a^*b^*$— may better synthesise realistic visuals of anomalous strawberries than on the RGB, originally designed in (Choi et al. 2021b). In fact, Fig. 3 shows that even the same random permutation of channels can result in highly different augmented images between colour spaces. This is because colour spaces encode a unique property of colour in each channel (cf. Fig. 2); for instance, while RGB keeps the chromaticity values of red, green, and blue, HSV, $YCbCr$, and $L^*a^*b^*$ each use some quantification of brightness in the channel of V, Y, and L, respectively, determining other unique properties of colour in the other two channels (Szeliski 2010) — e.g., hue in H, blue colour difference in Cb, and green-to-red colour in a^*.

In our experiments, we thus evaluate each colour space to discover the best visual domain for representation learning to identify anomalous strawberries from image data.

Metric	DCAE	ROT	CP	CH-R RGB	CH-R HSV	CH-R YCbCr	CH-R L*a*b*
ROC	.715 ±.002	.736 ±.005	.736 ±.007	.804 ±.014	.778 ±.007	.757 ±.001	**.810** ±.007
PR	.340 ±.003	.335 ±.016	.337 ±.006	.496 ±.022	.457 ±.030	.409 ±.015	**.547** ±.030

Table 1: Mean AUC-ROC and AUC-PR scores with standard deviations achieved by three separate runs of each method.

Curriculum Learning Inspired by (Bengio et al. 2009; Hacohen and Weinshall 2019), we build training procedures in which CH-Rand is regulated to generate gradually more difficult images in augmentation. In other words, although we initially consider all possible images for augmentation, CH-Rand is set to discard the channel sequences that would create relatively easy examples, as the training proceeds.

More formally, CH-Rand draws an arbitrary function $\pi_c : c \mapsto c' \in \chi'_c$ for reassignment of each channel c, where $\chi'_c \in \mathcal{P}(\chi) \backslash \mathcal{B}_c \backslash \varnothing$ as $\mathcal{B}_c$ can be *specified* to define a possible c' from χ'_c at certain times for curriculum; for example, if $\chi'_{L^*} = \{L^*\}$, $\chi'_{a^*} = \{a^*\}$, and $\chi'_{b^*} = \{L^*, a^*, b^*\}$, the producible domain can be limited by the randomness only in the b* channel. Therefore, Equation (1) is replaced by $a^c_{w,h} = i^{\pi_c(c)}_{w,h}$, and also note that we repeat drawing π_c until satisfying $\exists c \in \chi, c \neq \pi_c(c)$ to avoid $\mathcal{A} = \mathcal{I}$.

A more specific use case is introduced in the next section with the empirical results of performance improvement.

Experimental Results

In this section, we first present a description of Riseholme-2021 dataset, followed by the information of experimental setups, such as the architecture of deployed deep network and the protocols in training and evaluation. Evaluation results on each domain-control technique are then reported along with qualitative examination on challenging samples.

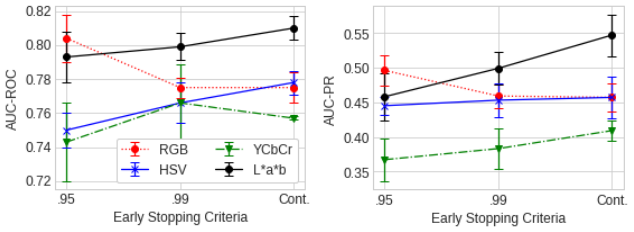
Dataset & Technical Details & Evaluation Criteria

Most settings presented in this study replicate the experimental design in (Choi et al. 2021b). Riseholme-2021 dataset[1] is used to train a neural network only with the images of normal strawberries — i.e., instances from the `ripe` (13.1%), `unripe` (68.4%), and `occluded` (14.2%) categories — to classify anomalous ones (4.3%) during test. For this OC scenario, we utilise the predefined exclusive set of `Train`, `Val`, and `Test` containing 70%, 10%, and 20% of normal samples, respectively, and all `Anomalous` examples are used only for test. Also, each image is resized to 64×64 pixels, to which traditional augmentations such as horizontal/vertical flips and colour jitter[2] are randomly applied in the RGB space. We then use OpenCV-Python[3] to change the colour space to execute CH-Rand on it, and finally, all pixel values are processed to be within $[-1, 1]$.

[1]https://github.com/ctyeong/Riseholme-2021

[2]https://pytorch.org/vision/stable/transforms.html

[3]https://docs.opencv.org/4.5.4/de/d25/imgproc_color_conversions.html



(a) (b)

Figure 4: Trends of (a) AUC-ROC and (b) AUC-PR scores in different colour models with various amounts of training. .95 and .99 are the early-stopped training when the validation accuracy reaches the corresponding number, whereas Cont. is persistent training for 1.5K epochs.

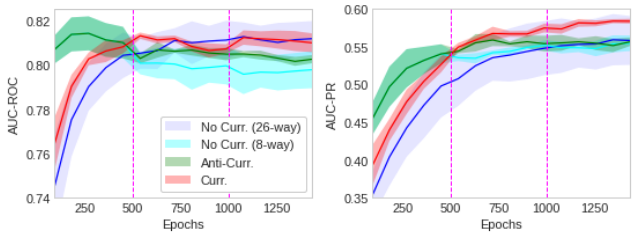
Moreover, we implement a deep network for OC based on the official code available online[4]. In particular, its structure is of 5 ConvLayers followed by 2 DenseLayers as the number of 3×3 convolutional filters increases by double at each layer — i.e., $64, 128, 256, 512$, and 512. Furthermore, a BatchNorm layer and a 2×2 MaxPool layer are applied after each ConvLayer, and the DenseLayers adopt 256 and 1 output nodes, respectively. Also, every node uses the LeakyReLU function for activation except in the last DenseLayer with a sigmoid function instead. Lastly, the outputs of the first DenseLayer are used as the learnt representations g_θ to gain anomaly scores for tested images based on the distance to the $k = 1$ nearest neighbor in training data $\mathcal{D}$.

For evaluation, we compute the Area Under the Curve (AUC) of both the Receiver Operating Characteristic (ROC) and Precision-Recall (PR). The former indicates the average rate of correct classification within each class (e.g., normal and anomalous), whereas the latter measures it only for minority class (anomalous) but additionally considers the *proportion* of majority-class (normal) samples in prediction of anomaly to better quantify performance in highly skewed distributions (Davis and Goadrich 2006). In particular, every reported score here is the mean from three individual models that each have reached the highest validation accuracy for 1.5K epochs unless mentioned otherwise.

Colour Space Effect

We here compare the performances obtained by different colour spaces against the state-of-the-art models run in (Choi et al. 2021b): Deep Convolutional Autoencoders and other SL approaches — e.g., RotNet (Gidaris, Singh, and Komodakis 2018) and CutPaste (Li et al. 2021). Specifically, Table 1 shows all baselines perform worse than any of CH-Rand-based methods. Still, the reliability of CH-Rand appears to depend highly on the selection of colour space because 33% improvement in AUC-PR is observed only with the change of applied colour scheme (e.g., $YCbCr \rightarrow L^*a^*b^*$). In particular, the model with $L^*a^*b^*$ reached the best performance among others especially leading to a sig-

[4]https://github.com/ctyeong/CH-Rand



(a) (b)

Figure 5: (a) AUC-ROC and (b) AUC-PR scores in $L^*a^*b^*$ models with and without curriculums over 1.5K training epochs. Magenta vertical lines indicate the timings where curriculums adjust the difficulty of task as designed. 8-way model only uses eight sequences in group L^* for CH-Rand.

nificantly higher AUC-PR than the originally suggested design with RGB. This performance margin actually implies that the $L^*a^*b^*$ space can effectively simulate the visual anomalies expected in strawberries to learn representations of a better quality.

This high relevance is explained also by Fig. 4, which reveals that with $L^*a^*b^*$, AUC-ROC and AUC-PR scores both keep increasing as training continues even after validation accuracy has reached .99. In contrast, other models such as RGB show consistent performance drops whilst the accuracy converges for the validation set. In fact, the pretext task with RGB images indicated relatively high relevance compared to other state-of-the-art approaches such as ROT and CP in (Choi et al. 2021b), but here we show that the utility of CH-Rand can be *enhanced* by using $L^*a^*b^*$ instead.

Curriculum Learning

We evaluate a learning curriculum with the colour space of $L^*a^*b^*$. To be specific, we first categorise 26 possible augmented sequences using the three channels into three groups, so-called group L^*, a^*, and b^*, based on the source information assigned to the first channel, L^*, in the resulting image — i.e., the output of $\pi_{L^*}(L^*)$. For instance, group a^* contains the nine sequences, such as $\{a^*L^*L^*, a^*L^*a^*, a^*L^*b^*, ..., a^*b^*b^*\}$.

Our curriculum is designed to involve all groups in augmentation initially but start to *ignore* group b^* and a^* from 500th and $1,000$th epoch, respectively, to set the highest difficulty of classification at last epochs solely with group L^*. This approach is based on the intuition that when an image loses the original values of its first channel, the overall appearance can dramatically change, and thus, the group a^* and b^* would be of relatively easy examples to distinguish.

We also test two more relevant models:

- 8-*way w/* L^*: Only use eight sequences in group L^*.
- *Anti-Curriculum*: Start with group b^*, and also consider group a^* and L^* after 500 and 1K epochs, respectively.

Table 2 shows a notable improvement with the curriculum method in AUC-PR — i.e., $> 16\%$ and $> 5\%$ over

Metric	No Curr. 26-way w/ All	No Curr. 8-way w/ L^*	Anti-Curr.	Curr.
ROC	.810 ±.007	.796 ±.009	.802 ±.002	**.811** ±.004
PR	.547 ±.030	.550 ±.009	.556 ±.005	**.576** ±.006

Table 2: Performance of $L^*a^*b^*$ models with and without curriculums. 8-way with L^* is the case only using eight sequences in group L^* for CH-Rand.

the RGB model and the $L^*a^*b^*$ without curriculum, respecitvely — even though the same networks and datasets are utilised. Also, the lower performance in the 8-way model implies the importance of easy start with a large set of possible permutations as in our curriculum to finally convergence to a useful representation space. The suboptimal result from the anti-curriculum approach also supports this observation.

Furthermore, Fig. 5 explains that the curriculum generally *accelerates* learning of useful representations with a considerably lower variation in both metrics than naïve training. In contrast, learning with hard samples first in 8-way and anti-curriculum methods tends to degrade the quality of representations albeit their initial performance can seem promising.

Challenging Cases

Figure S2 depicts samples most challenging for the $L^*a^*b^*$ model with curriculum. Specifically, normal instances are found to be difficult to classify when fruits are occluded, or brightness is too high or low. Also, red diseased berries are shown to be confusing along with anomalies in early developmental stages before strawberries are fully formed on flower buds probably due to some extent of visual similarity to healthy ripe and unripe strawberries, respectively.

Conclusion & Future Work

We have investigated Channel Randomisation (CH-Rand) augmentation with novel domain-control strategies to learn useful representations of visual anomalies in strawberries whilst a neural network classifies the augmented images. In particular, we have discovered that the constructed domains of visual data in the $L^*a^*b^*$ colour space can best guide neural networks to learn informative representations among other colour expression models. Moreover, curriculum learning could further improve obtained representations by regulating available channel sequences for augmentation to gradually increase the difficulty of pretext task. With the optimal setup, the trained model has led to a $> 16\%$ higher AUC-PR than the previous best on Riseholme-2021 dataset.

For future work, we plan to address the discussed issues with extreme conditions to perform more robust predictions. Furthermore, we could attempt to integrate with object detectors to build a more realistic pipeline in which the inputs are not necessarily images cropped around fruits. Also, more sophisticated curriculums could be invented to estimate and make use of image-level difficulties to further maximise learning effect.

References

Bengio, Y.; Louradour, J.; Collobert, R.; and Weston, J. 2009. Curriculum learning. In *Proceedings of the 26th annual international conference on machine learning*, 41–48.

Bergmann, P.; Batzner, K.; Fauser, M.; Sattlegger, D.; and Steger, C. 2021. The MVTec anomaly detection dataset: a comprehensive real-world dataset for unsupervised anomaly detection. *International Journal of Computer Vision*, 129(4): 1038–1059.

Choi, T.; Pyenson, B.; Liebig, J.; and Pavlic, T. P. 2021a. Identification of Abnormal States in Videos of Ants Undergoing Social Phase Change. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, 15286–15292.

Choi, T.; Would, O.; Salazar-Gomez, A.; and Cielniak, G. 2021b. Self-supervised Representation Learning for Reliable Robotic Monitoring of Fruit Anomalies. *arXiv preprint arXiv:2109.10135*.

Davis, J.; and Goadrich, M. 2006. The relationship between Precision-Recall and ROC curves. In *Proceedings of the 23rd international conference on Machine learning*, 233–240.

Gidaris, S.; Singh, P.; and Komodakis, N. 2018. Unsupervised Representation Learning by Predicting Image Rotations. In *International Conference on Learning Representations*.

Hacohen, G.; and Weinshall, D. 2019. On the power of curriculum learning in training deep networks. In *International Conference on Machine Learning*, 2535–2544. PMLR.

Hendrycks, D.; Mazeika, M.; Kadavath, S.; and Song, D. 2019. Using Self-Supervised Learning Can Improve Model Robustness and Uncertainty. *Advances in Neural Information Processing Systems*, 32: 15663–15674.

Li, C.-L.; Sohn, K.; Yoon, J.; and Pfister, T. 2021. CutPaste: Self-Supervised Learning for Anomaly Detection and Localization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 9664–9674.

Perera, P.; and Patel, V. M. 2019. Learning deep features for one-class classification. *IEEE Transactions on Image Processing*, 28(11): 5450–5463.

Szeliski, R. 2010. *Computer vision: algorithms and applications*. Springer Science & Business Media.

Yi, J.; and Yoon, S. 2020. Patch SVDD: Patch-level SVDD for anomaly detection and segmentation. In *Proceedings of the Asian Conference on Computer Vision*.

A CH-Rand Examples on L*a*b*

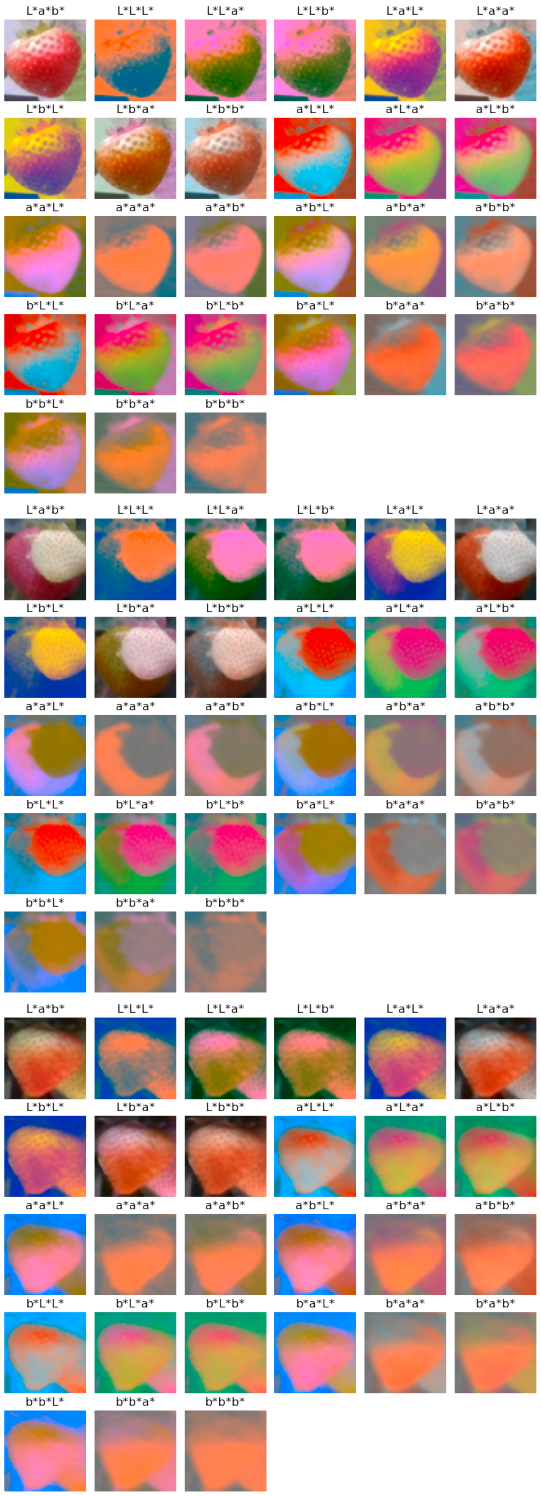


Figure S1: All possible *channel-randomised* outputs for three individual instances. In each case, the top-left displays the original input, followed by 26 augmentations of group L^*, a^*, and b^* in order. The label above each image shows the source channels mapped to the new L^*, a^*, and b^*

B Challenging Examples

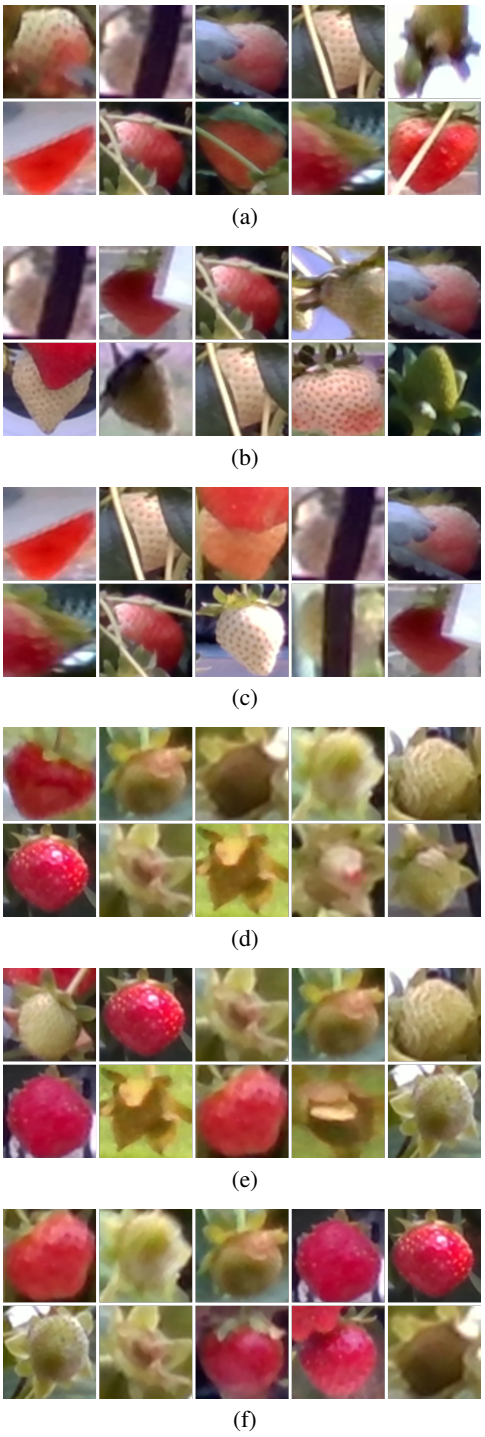


Figure S2: Strawberry images most challenging for three independent $L^*a^*b^*$ models trained with curriculum. For each model, 10 images of normal strawberries with highest anomaly scores are displayed in (a)–(c), and 10 images of anomalous ones with lowest anomaly scores in (d)–(f).