Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Where to Put Low-Rank Capacity? A Grouped-Split Study of LoRA for Industrial Defect Segmentation

Shaoxuan Zhang

2023113033@stu.hit.edu.cn

Harbin Institute of Technology

Research Article

Keywords: surface defect segmentation, SegFormer, LoRA, parameter-efficient fine-tuning, grouped-split evaluation, magnetic tile

Posted Date: November 1st, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11517758/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Surface defects on magnetic tiles and steel strips are small, sparse, and visually heterogeneous. Full fine-tuning of a segmentation transformer can learn these patterns, but it is costly and data-hungry. We ask where low-rank adaptation should put its capacity, and how the result changes when whole filename groups are held out. We freeze a SegFormer-B0 encoder and adapt attention projections and MLP layers with LoRA. On the Magnetic Tile Defect dataset, using six filename groups as held-out test sets and three seeds, q/k/v/o plus MLP reaches 0.779 defect IoU, against 0.723 for attention-only q/k/v/o and 0.772 for full fine-tuning. The difference from Full is 0.7 points, with a 95% interval from − 2.0 to + 3.4 ($p = 0.54$); the study does not establish equivalence. Attention-only LoRA already supplies most of the adaptation gain, and the MLP adapters close the residual gap on this dataset. On NEU-Seg, a steel-surface dataset, the same configuration reaches 0.699, against 0.684 for attention-only LoRA and 0.715 for Full; the MLP gain is smaller and not significant. The configuration trains 0.543M parameters, of which 0.395M is the decoder head and 0.148M is adapter-specific. A BitFit baseline reaches 0.608, so the gain is not explained by a few bias terms. A bottleneck adapter is close to qkvo + MLP on both datasets. Random image splits compress the gap between attention-only LoRA and Full, while group-held-out splits expose a larger gap; we report both and treat the difference as split sensitivity rather than proof of leakage.

1. Introduction

Visual inspection remains a bottleneck in magnetic-tile and steel-strip manufacturing. A blowhole, a hairline crack, or a patch of fraying may occupy only a few dozen pixels, and the surrounding texture is often louder than the defect itself. Segmentation models can help, and hierarchical transformers such as SegFormer [1] have made the task more practical. The catch is familiar: these models are pre-trained on natural images, and adapting them to grayscale industrial images requires labeled data, time, and GPU memory.

Full fine-tuning is the default answer. It works, but it is a heavy hammer for a small factory dataset. Parameter-efficient fine-tuning (PEFT) offers a lighter option: keep the pre-trained encoder frozen, add a small number of trainable parameters, and adapt the representation to the target material. LoRA [2] is attractive because it adds a low-rank update to an existing projection and leaves the original weights untouched. Two practical questions remain open. First, where should the low-rank update be inserted in a segmentation transformer? Second, does the adapter still help when a whole filename group is held out, rather than a random subset of the same images?

Our first random-split result looked encouraging and slightly suspicious. Attention-only LoRA appeared almost identical to full fine-tuning. When we held out an entire filename group, the gap widened. A second change followed: the attention target that looked decisive on the random split lost most of its advantage once the split changed. What survived was the MLP adapter. The paper keeps both protocols and uses the group-held-out result as the main evidence. We use the term filename group rather than

acquisition group because the public dataset documentation does not define what the expN prefix represents; the split is group-aware at the filename level, and we do not claim strip-level separation.

This paper makes four contributions. First, we evaluate PEFT under a grouped filename split on MTD, so that no group appears in both train and test. Second, we show that attention-only q/k/v/o already supplies most of the adaptation gain, while MLP adapters close the residual gap: 0.723 for attention-only LoRA, 0.779 for qkvo + MLP, and 0.772 for Full. Third, we repeat the comparison on NEU-Seg and obtain a similar mean ordering of PEFT methods, although the MLP gain is small and not significant, and full fine-tuning remains 1.6 points ahead. Fourth, we show that random image splits compress the LoRA-to-Full gap relative to a group-held-out split, and we report that split sensitivity without attributing it to a single mechanism. The result is a bounded empirical study, not a new adapter architecture.

2. Related Work

2.1 Surface defect segmentation

Industrial surface inspection has a long history in computer vision [12]. U-Net [3] and DeepLabv3+ [4] remain strong supervised baselines, and datasets such as MVTec AD [5] and KolektorSDD [6] have made evaluation more reproducible. Magnetic-tile inspection has moved from texture-feature clustering [13] and handcrafted surface descriptors [14] to CNN detectors such as ELCNN [15] and local-structure methods [16]. Steel-surface segmentation has been studied with multi-branch U-Nets [17] and domain-adaptation detectors [18]. SegFormer [1] combines a hierarchical transformer encoder with a lightweight MLP decoder; its encoder is a natural candidate for transfer learning, but full adaptation is expensive on small datasets. Magnetic tile and steel strip are two materials with different textures and defect morphologies, which makes them a useful pair for testing whether a PEFT result generalizes beyond one dataset.

2.2 Parameter-efficient fine-tuning

PEFT methods adapt a frozen model with a small number of new parameters. Adapters [7] insert bottleneck modules into transformer blocks; AdapterFusion composes adapter modules across tasks [33], and Compacter reduces their parameter cost [34]; BitFit [8] trains only bias terms; LoRA [2] decomposes a weight update into two low-rank matrices. LoRA has been widely used for language models and is increasingly used in vision transformers [9]. Vision-oriented variants include visual prompt tuning [19], AdaptFormer [20], and convolutional bypass adapters [21]; a unified view relates these families [22], and LoRA has been extended to few-shot segmentation [23]. A recent guide surveys the trade-offs among these approaches [24]. The original LoRA study already includes an MLP-target ablation, so adapting fc1/fc2 is not a new adapter design. What is less settled is how that choice behaves in dense prediction, on industrial images, and under a grouped split. Our comparison is positioned there, not as a new PEFT method.

Most PEFT papers report a single target set and a single split. That design makes it hard to separate the effect of the adapter from the effect of the evaluation protocol. We vary both. The target sets range from q/v to q/k/v/o to q/k/v/o plus MLP, and the protocols range from a random image split to a six-group rotation. The contribution is the comparison itself; we do not propose a new adapter architecture. The comparison matters because the two design choices interact. The target set that looks best on a random split can lose its advantage once a whole group is held out.

2.3 Evaluation protocols

Random image splits are common in industrial defect papers. When images share a filename group, random splits can place related crops on both sides; shortcut learning explains why this can inflate performance [25], and such leakage is a recognized reproducibility problem [26]. Group-held-out splits are stricter at the filename level and are the main protocol here. Grouped and domain-generalization protocols are one response to this failure mode [27, 28]. We report the random-split result separately because the contrast is informative, but we treat it as split sensitivity rather than measured leakage.

3. Methods

3.1 Problem and backbone

Given an RGB image x of size H×W×3, we predict a binary mask y of size H×W, where 1 marks a defect. Images and masks are resized to 384×384 for MTD and 256×256 for NEU-Seg. SegFormer-B0 [1] is the backbone: its encoder produces multi-scale features, and its decoder fuses them into a mask. The encoder is pre-trained on ADE20K and kept frozen in all PEFT experiments.

3.2 Low-rank adaptation in attention and MLP

For a frozen linear projection W of size d×k, LoRA learns W' = W + BA, where B is d×r, A is r×k, and r is much smaller than min(d,k). We compare three families: q/v, q/k/v/o, and the same attention targets plus the two MLP projections (fc1 and fc2). Rank is varied over 4, 8, and 16. The original W stays frozen; only A, B, and the decoder head receive gradients. Figure 1 shows the pipeline. The default configuration in the main experiments is q/k/v/o plus MLP with r = 8, alpha = 16, and dropout 0.1; it trains 0.543M parameters.

Most of the budget sits in the decoder head. That head has about 0.395M parameters; q/k/v/o at rank 8 adds 0.066M, and the two MLP projections add 0.082M. The comparison in Table 3 is therefore close to a like-for-like test of where the capacity goes, and a larger rank shows little accuracy return.

3.3 Objective and optimization

Defect pixels are rare. Class imbalance is a well-studied problem in dense prediction [29, 35]. We use a class-weighted cross-entropy plus Dice loss [30, 32], with a positive weight estimated from the training pixel ratio and capped at 20; the weighting follows the spirit of focal loss [31] while keeping the standard

cross-entropy form. Dice is computed on the foreground logits. MTD models use 384×384 inputs, batch size 2, 12 epochs, and AdamW; NEU-Seg models use 256×256, batch size 2, and 8 epochs. The decoder head uses lr = 1e-3, LoRA uses lr = 5e-4, and weight decay is 1e-4. Full fine-tuning uses lr = 6e-5. Training runs on a single 8GB RTX 4060 laptop GPU with PyTorch, HuggingFace Transformers, and PEFT. No test-time adaptation is used.

3.4 Grouped filename splits

MTD filenames start with expN. The public dataset README (accessed 2026-10-27) does not define what expN represents, so we treat each expN as a filename group rather than an acquisition unit. We run leave-one-group-out (LOGO): each of exp1–exp6 is used once as test, the next group is validation, and the remaining groups plus exp0 form the training set. No group appears in two splits. We repeat every configuration with three seeds. This is a group-aware split at the filename level; we do not claim that it separates strips or acquisition runs. For NEU-Seg, we use the official training/test split, hold out 10% of training for validation, and report the official test set. The NEU-Seg split is weaker than LOGO, but it provides an independent material and label source. The exact fold manifests are provided with this work at https://github.com/ZhangShaoXuan-HIT/mtd-lora-seg (tag v1.0.0).

3.5 Baselines

We compare five baselines. Frozen trains only the decoder head. BitFit trains all bias terms plus the decoder head. LoRA q/v and LoRA q/k/v/o train the corresponding low-rank matrices plus the decoder head. qkvo + MLP adds the two MLP projections. Full fine-tuning updates all parameters. All methods use the same loss, augmentation, epochs, and seeds. The MTD group-split comparison uses the exp6 group as test and exp3 as validation; this is also the group used to choose the target set and rank. The LOGO comparison rotates all six groups. A pretrained DeepLabV3 model was run under the same budget but is excluded from the comparison because it did not converge.

4. Experimental Setup

4.1 Datasets

MTD [10] contains 1,344 grayscale images from a magnetic-tile production line: 115 blowhole, 85 break, 57 crack, 32 fray, 952 free, and 103 uneven. Each image has a pixel-level mask; we threshold the mask at 127 to obtain binary labels. The dataset is highly imbalanced: free tiles dominate, and fray has only 32 images. The exp6 test group contains 225 images: 19 blowhole, 15 break, 10 crack, 6 fray, 158 free, and 17 uneven.

NEU-Seg [11] contains 4,470 steel-surface images with binary masks: 3,630 training and 840 test. We hold out 10% of the training set for validation, leaving 3,267 training and 363 validation images. Unlike MTD, NEU-Seg has no defect-free images in the official training/test folders; every image contains a defect. This makes the task a pure foreground segmentation problem and provides a useful contrast to the imbalanced MTD setting.

4.2 Metrics and implementation

We report foreground defect IoU and Dice over the whole test set. Both metrics accumulate intersections and unions over all images, rather than averaging per-image scores. The main MTD result is the mean over six groups and three seeds; the main NEU-Seg result is the mean over three seeds. Full fine-tuning on MTD LOGO was run with three seeds, and all methods use the same six groups.

5. Results

5.1 MTD: leave-one-group-out

Table 1 and Fig. 2 summarize the six-group rotation. Frozen is stable at 0.482 defect IoU; attention-only LoRA reaches 0.723; adding MLP adapters lifts the mean to 0.779. Full fine-tuning reaches 0.772. The MLP configuration is therefore 5.65 points above attention-only LoRA (paired t-test over six group means, n = 6, df = 5, p = 0.0025) and 0.70 points above Full; that difference is not significant (p = 0.54), and the 95% interval is -2.0 to + 3.4. It trains 0.543M parameters, of which 0.395M is the decoder head and 0.148M is adapter-specific. Every group improves over attention-only LoRA; the gain ranges from 1.3 points on exp2 to 7.8 points on exp1. The MLP configuration is above Full on four of six groups, below by 3.8 points on exp5 and by 0.3 points on exp6.

Among the adapted methods, exp4 is hardest and exp6 is easiest; Frozen is least accurate on exp5, and Full peaks on exp5. qkvo + MLP wins on all six groups relative to attention-only LoRA.

Table 1
MTD LOGO defect IoU (3 seeds); group means are reported in Table 7.

Test group	Frozen	LoRA qkvo	qkvo + MLP	Full
exp1	0.4768	0.7158	0.7938	0.7902
exp2	0.5226	0.7297	0.7422	0.7190
exp3	0.4688	0.7055	0.7733	0.7387
exp4	0.5172	0.6568	0.7028	0.6816
exp5	0.4438	0.7517	0.8277	0.8655
exp6	0.4653	0.7781	0.8366	0.8393

5.2 NEU-Seg: steel surface

The same ordering transfers to steel. On NEU-Seg, Frozen reaches 0.483 defect IoU and 0.651 Dice; attention-only LoRA reaches 0.684 and 0.812; qkvo + MLP reaches 0.699 and 0.823; Full reaches 0.715 and 0.834. MLP adaptation improves over attention-only LoRA by 1.57 points (paired t-test over three

seeds, n = 3, df = 2, p = 0.235, not significant) and remains 1.56 points below Full (p = 0.013). The three MLP seeds are tight (standard deviation 0.35 points). The mean gain is positive in two of three seeds; seed 0 is -0.2 points. This is a mean-level directional agreement, not a uniform ordering across seeds.

Steel is the harder test for our interpretation. The MLP gain is real but modest, and Full keeps a clear lead. One reading is that steel defects are larger and more contrastive, so attention adaptation already captures much of the signal and the MLP has less room to help. Another reading is that the NEU-Seg split is closer to the training distribution, which compresses differences among PEFT methods. We cannot separate these explanations with the current data. The honest statement is that MLP adaptation is the best PEFT on both datasets, and that its advantage over Full is material only on MTD.

Table 2
NEU-Seg results over three seeds (mean ± std). MLP adaptation is the best PEFT, but Full remains 1.6 points ahead.

Method	defect IoU	Dice
Frozen baseline	0.4826 ± 0.0176	0.6509 ± 0.0160
LoRA qkvo	0.6835 ± 0.0109	0.8120 ± 0.0077
qkvo + MLP	0.6992 ± 0.0035	0.8230 ± 0.0024
Bottleneck adapter	0.6879 ± 0.0140	0.8150 ± 0.0099
Full fine-tune	0.7148 ± 0.0014	0.8336 ± 0.0009

5.3 Baselines and parameter efficiency

Table 3 isolates the effect of the adaptation method on the MTD exp6 design-selection group. This group split uses exp3 for validation and all remaining groups for training, so its exp6 values differ from the exp6 LOGO fold in Table 1. BitFit improves over Frozen by 13.7 points, which confirms that adapting the encoder helps. Attention-only LoRA improves over BitFit by another 13.6 points, so the gain is not explained by a few bias terms. Adding MLP adapters lifts the result to 0.852 defect IoU and 0.920 Dice, 10.8 points above attention-only LoRA and 3.4 points above Full. The bottleneck adapter lands at 0.840, close to qkvo + MLP and also above Full. This is the one place where our attention-only expectation broke, and we report it openly. On NEU-Seg, MLP adaptation is still the best PEFT, but Full remains 1.6 points ahead. A scratch U-Net and a pretrained DeepLabV3 were run under the same budget and did not converge to a useful result; we exclude both from the comparison rather than presenting a training failure as an architecture result. Figure 3 places the methods on an accuracy-versus-parameters plane; qkvo + MLP and the adapter sit above Full, and their uncertainty ranges overlap.

Table 3

MTD exp6 design-selection group: adaptation method vs trainable footprint (3 seeds). This group was used to choose the target set and rank, so it is not an unseen-test estimate. qkvo + MLP is the strongest method here; its advantage over Full does not transfer to NEU-Seg.

Method	Trainable	defect IoU	Dice
Frozen baseline	0.40M	0.4709 ± 0.0104	0.6402 ± 0.0097
BitFit	0.41M	0.6082 ± 0.0156	0.7563 ± 0.0121
LoRA qkvo	0.46M	0.7438 ± 0.0320	0.8527 ± 0.0209
qkvo + MLP	0.54M	0.8516 ± 0.0113	0.9198 ± 0.0066
Bottleneck adapter	0.49M	0.8404 ± 0.0071	0.9133 ± 0.0042
Full fine-tune	3.71M	0.8174 ± 0.0161	0.8994 ± 0.0098

5.4 Per-class behavior

The aggregate gain is not uniform. Table 4 breaks down the exp6 design-selection group by defect class for Frozen and attention-only LoRA. Crack remains the hardest class for both models, although LoRA improves it from 0.384 to 0.541. Uneven benefits most (+ 0.300), while Fray is already high under Frozen and gains less (+ 0.070). The mean over the five defect classes is 0.546 for Frozen and 0.720 for LoRA; this unweighted mean is not the pooled foreground IoU in Table 3. Per-class results for qkvo + MLP are not available, which is a limitation of this analysis.

Table 4

Per-class IoU on the MTD exp6 design-selection group (seed 0), Frozen vs attention-only LoRA; qkvo + MLP is not available. The mean is an unweighted average over five defect classes.

Class	Frozen IoU	LoRA IoU	Δ
Blowhole	0.5431	0.6829	+ 0.1398
Break	0.5629	0.7691	+ 0.2062
Crack	0.3840	0.5407	+ 0.1567
Fray	0.7899	0.8595	+ 0.0696
Uneven	0.4488	0.7486	+ 0.2998
Mean (5 classes)	0.5457	0.7202	+ 0.1745

5.5 Qualitative comparison

Figure 4 shows four defects from the exp6 design-selection group. The frozen model often produces a loose blob around the defect; attention-only LoRA follows thin cracks and frayed edges more closely. The

difference is largest on Crack, where the frozen prediction spills into the surrounding texture, and smallest on Fray, where the defect region is large and both models find it. The MLP variant is quantitatively stronger, but the examples show only the attention contrast; they are illustrative and not evidence for the final recipe.

5.6 Why the evaluation protocol matters

Table 5 compares three protocols on the same attention-only model family. Under a random image split, LoRA reaches 0.759 and Full 0.777. Under the exp6 group split, LoRA is 0.744 and Full 0.817. Across the six-group LOGO mean, LoRA is 0.723 and Full 0.772. The random split changes train/test composition and may allow related crops to cross the split. We do not have a duplicate or near-duplicate measurement, so we treat the difference as split sensitivity rather than proof of leakage. Figure 5 shows the two constructions: in the left panel, every group contributes to both train and test; in the right panel, exp6 is kept whole for testing. The group-held-out result is harder, and we use it as the main evidence.

Table 5
Protocol comparison for attention-only LoRA (3 seeds). Random splitting compresses the gap to Full relative to group-held-out evaluation. The three rows also differ in train/test composition, so the contrast is descriptive.

Protocol	Frozen	LoRA qkvo (attn only)	Full
Random images	0.4415	0.7590	0.7774
Group exp6	0.4709	0.7438	0.8174
LOGO mean (6 groups)	0.4824	0.7229	0.7724

5.7 Design ablation

Table 6 reports the group-split ablation over three seeds. Without MLP adapters, q/v and q/k/v/o are almost tied (0.744 vs 0.746 at rank 16/8); the earlier random-split result that favored q/k/v/o by seven points does not survive. Adding MLP adapters is the dominant effect: q/v jumps to 0.811 and q/k/v/o to 0.852. With MLP adapters present, q/k/v/o is still 4.1 points above q/v, so the target set is a secondary but real factor. Rank is weak: r8 is slightly better than r16 for q/k/v/o + MLP (0.852 vs 0.842) and uses fewer parameters. We therefore used q/k/v/o + MLP with rank 8 in the main experiments, while treating the target and rank comparison as exploratory.

Table 6
Single-group design probe (test=exp6, 3 seeds). This table was used to choose the configuration and is not an unseen-test estimate; the rank and target-set differences are exploratory.

Configuration	defect IoU
r4 q/v	0.692
r8 q/v	0.720
r16 q/v	0.744
r8 q/k/v/o	0.746
r8 q/v + MLP	0.811
r8 q/k/v/o + MLP	0.852
r16 q/k/v/o + MLP	0.842

5.8 Cross-dataset summary

Table 7 collects the main results. On MTD LOGO, qkvo + MLP is 0.70 points above Full, with a 95% interval from − 2.0 to + 3.4 (p = 0.54); we do not claim equivalence. Because exp6 was used to choose the configuration, we also computed the five-group mean excluding exp6: qkvo + MLP 0.768, Full 0.759, a difference of + 0.89 (p = 0.52). Removing exp6 does not change the substantive conclusion. On NEU-Seg, qkvo + MLP is 1.56 points below Full. The exp6 design group is not included in Table 7. The two datasets have different class compositions and input sizes, so their point differences should not be compared directly.

Table 7
Main results across datasets. Δ is qkvo + MLP minus Full in defect IoU points. The MTD difference is not an equivalence result; the 95% interval is -2.0 to + 3.4.

Dataset / protocol	Frozen	LoRA qkvo	qkvo + MLP	Full	Δ (points)
MTD LOGO (6 groups)	0.4824	0.7229	0.7794	0.7724	+ 0.70
NEU-Seg (steel)	0.4826	0.6835	0.6992	0.7148	-1.56

6. Discussion

Why does the frozen decoder baseline fail so badly? The ADE20K features are rich, but they were not trained to separate faint grayscale textures. The decoder head can only reweight what the encoder already exposes. Attention-only LoRA changes the attention pattern, but it leaves the MLP layers untouched; on MTD it reaches 0.723. Adding low-rank updates to the MLP projections lifts the same

model to 0.779. The MLP layers evidently hold part of the capacity needed to separate these textures. Full fine-tuning remains ahead on NEU-Seg by 1.6 points. On MTD the two methods are not significantly different, but the interval is wide enough that we do not call them equivalent.

The design probe suggests a narrower lesson. Attention target selection is a secondary knob: without MLP adapters, q/v and q/k/v/o are almost tied (0.744 vs 0.746); with MLP adapters, q/k/v/o is 4.1 points above q/v (0.852 vs 0.811). Rank is weak: r8 is slightly better than r16 for q/k/v/o + MLP and uses fewer parameters. These differences come from one design group and should be read as exploratory. The main change is whether the MLP projections receive low-rank capacity at all; attention-only LoRA alone is consistently weaker.

The adapter result points in the same direction. A bottleneck adapter is another way to add capacity around the MLP path; it reaches 0.776 on MTD LOGO, close to qkvo + MLP (0.779, $p = 0.71$), and the two are not significantly different. On NEU-Seg, qkvo + MLP is the best PEFT at 0.699, but Full remains 1.6 points ahead. The practical recommendation is narrower than a general recipe: adding MLP capacity is a reasonable option when attention-only LoRA is insufficient, but it should be validated on a held-out group.

The protocol finding is a warning about split sensitivity. The random-split result looked more favorable than the group-held-out result. We keep both because the difference is informative. Under MTD LOGO, attention-only LoRA is 5.0 points behind Full, and adding MLP adapters closes that gap. The comparison is descriptive: the two splits also differ in train/test composition, and we have not measured near-duplicate overlap.

Limitations. The expN prefix is not documented as an acquisition session; the split is at filename level and is not verified at strip or batch level. Full fine-tuning on MTD LOGO uses three seeds, but its group-to-group spread is larger than that of qkvo + MLP (0.065 vs 0.047, population standard deviation over the six group means), so the mean comparison should be read with that heterogeneity in mind. NEU-Seg has a random train/validation split and no defect-free images. Both tasks are binary; multi-class defect segmentation is left for future work. The MTD exp6 result (0.852) comes from the group used to choose the configuration and is not an unseen-test estimate. A scratch U-Net and a pretrained DeepLabV3 did not converge within the same budget; both are documented as non-converged runs rather than as general statements about the architectures. The p-values describe this set of groups and seeds, not a population of factories. The 8GB laptop GPU limits image resolution and batch size, and we do not test edge deployment latency.

6.1 Practical guidelines

For a practitioner with a small defect dataset, we would suggest four steps. First, establish a frozen decoder baseline; it is cheap and tells you how much the pre-trained features already know. Second, compare attention-only LoRA with a variant that also adapts the MLP projections; the MLP addition closed most of the residual gap on MTD in our experiments, but it is not a universal win. Third, use rank 8

as a default and spend the tuning budget on augmentation and loss weighting instead of chasing a larger rank. Fourth, evaluate on a held-out filename group if the data has group structure; a random split can make the adapter look better than it will be in deployment.

These guidelines are deliberately modest. LoRA does not remove the need for labels, and it does not beat full fine-tuning on NEU-Seg. It reduces the trainable-parameter budget while reaching comparable accuracy, which is useful when the alternative is a frozen baseline or no adaptation at all.

6.2 Threats to validity

The main internal threat is the group-level variation on MTD. Full has a larger spread across group means than qkvo + MLP (0.065 vs 0.047), and a different set of six groups could shift the mean and the ranking. The main external threat is the two-dataset scope: magnetic tile and steel are both grayscale industrial surfaces, and the conclusion may not transfer to color, textured, or three-dimensional surfaces. NEU-Seg also lacks defect-free images, so its task is easier to frame as foreground segmentation than MTD's imbalanced binary task. Construct validity is limited by binary labels; a multi-class evaluation could change the ranking of adapter targets. Finally, the 8GB GPU constrains resolution and batch size, and we have not measured latency on an edge device.

6.3 Why the MLP path may matter

Attention and MLP layers do different jobs in a transformer block. Attention decides which tokens exchange information; the MLP applies a channel-wise nonlinear transform to each token. A defect is often a subtle change in local texture and contrast, and that change must survive the channel mixing before the decoder can separate it from the background. A low-rank update in the MLP can act as a small, domain-specific recalibration of those channels. A low-rank update in attention changes routing, which may be less direct when the defect is defined by appearance rather than by position. This account is consistent with the adapter result: a bottleneck placed around the block output, after the MLP, reaches the same MTD level as qkvo + MLP. This remains an interpretation; the mechanism is not proven, and attention-map or feature-similarity analyses would be needed to test it.

6.4 What would change our conclusion

Three results would weaken the claim. A third material with color or strong specular reflection could reverse the ordering, especially if attention adaptation becomes more important than channel recalibration. A different six-group set could move the MTD means; the current gap is 0.7 points and not significant, but the interval still includes a two-point deficit, so the ordering is unresolved rather than equivalent. A multi-class evaluation could change the target-set ranking, because rare classes place different demands on channel capacity. We would also revise the claim if a held-out group on NEU-Seg reversed the MLP gain. The current evidence supports a bounded recommendation, and these tests define the boundary.

6.5 What the numbers do not say

Parameter count is not the same as wall-clock savings. LoRA still runs a forward and backward pass through the frozen encoder, so training time is close to full fine-tuning on this GPU. Inference cost is unchanged because the low-rank matrices can be merged into the frozen weights. The 0.543M trainable total contains 0.395M for the decoder head; the adapter-specific part is about 0.148M. A per-material package still needs the frozen encoder and the head, so its deployed footprint is roughly 3.86M parameters, slightly larger than the 3.71M full model. The benefit is a shared backbone with small per-material deltas, not a smaller single-model deployment. We did not measure peak memory or latency, and we do not claim either.

6.6 Broader implications

The result matters most for factories with several production lines and modest GPU budgets. A frozen backbone plus a small adapter can be trained per material and stored as a per-material delta, while the backbone is shared. That fits a maintenance workflow in which a new filename group arrives and the model must be updated quickly, but we have not measured the operational cost of that workflow. A random split can make a method look deployment-ready when a group-held-out test shows a gap. We would encourage reporting both, especially when the data has group structure. The approach does not replace full fine-tuning when accuracy is the only criterion. For magnetic tile, the parameter trade is reasonable; for steel, full fine-tuning remains the safer choice at this data scale.

7. Conclusion

We studied where to put low-rank capacity in a SegFormer encoder under a grouped filename split. Across six magnetic-tile groups and one steel-surface dataset, adding MLP adapters to q/k/v/o raised LoRA from 0.723 to 0.779 on MTD LOGO, and from 0.684 to 0.699 on NEU-Seg. On MTD the difference from Full is 0.70 points, with a 95% interval from − 2.0 to + 3.4 ($p = 0.54$); the study does not establish equivalence. On NEU-Seg qkvo + MLP remains 1.6 points below Full (0.715). The configuration trains 0.543M parameters, of which 0.148M is adapter-specific. BitFit captures only part of the gain, attention target selection is secondary, and rank is a weak knob. A bottleneck adapter is close to qkvo + MLP on MTD. Attention-only LoRA already supplies most of the adaptation gain; MLP adapters close the residual gap on MTD and about half of it on NEU-Seg. The practical recommendation is to compare both variants and validate on a held-out filename group. Future work will add stronger traditional baselines, multi-class segmentation, and a session-structured dataset with documented provenance.

Declarations

Competing interests

The author declares no competing interests.

Ethics approval

Not applicable. This study uses public datasets and does not involve human participants or animals.

Consent to participate

and publish

MTD and NEU-Seg are publicly available

[10, 11]. The exact split manifests are provided with this work at https://github.com/ZhangShaoXuan-HIT/mtd-lora-seg (tag v1.0.0). The dataset licences and access conditions follow the original repositories.

Funding

Author Contribution

Z conceived the study, implemented the models, ran the experiments, analysed the results, and wrote the manuscript.

Acknowledgement

The author thanks the creators of the Magnetic Tile Defect and NEU-Seg datasets for making their data publicly available.

Data Availability

MTD and NEU-Seg are publicly available. The exact split manifests are provided with this work at https://github.com/ZhangShaoXuan-HIT/mtd-lora-seg (tag v1.0.0). The dataset licences and access conditions follow the original repositories.

Code availability

References

1. Xie, E., Wang, W., Yu, Z., Anandkumar, A., Alvarez, J.M., Luo, P.: SegFormer: simple and efficient design for semantic segmentation with transformers. In: Advances in Neural Information Processing Systems (NeurIPS) (2021)

2. Hu, E.J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L., Chen, W.: LoRA: low-rank adaptation of large language models. In: International Conference on Learning Representations (ICLR) (2022)
3. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention (MICCAI), pp. 234–241 (2015)
4. Chen, L.C., Zhu, Y., Papandreou, G., Schroff, F., Adam, H.: Encoder-decoder with atrous separable convolution for semantic image segmentation. In: European Conference on Computer Vision (ECCV), pp. 801–818 (2018)
5. Bergmann, P., Fauser, M., Sattlegger, D., Steger, C.: MVTec AD: a comprehensive real-world dataset for unsupervised anomaly detection. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 9592–9600 (2019)
6. Tabernik, D., Šela, S., Skvarč, J., Skočaj, D.: Segmentation-based deep-learning approach for surface-defect detection. J. Intell. Manuf. **31**, 759–776 (2020). https://doi.org/10.1007/s10845-019-01476-x
7. Houlsby, N., Giurgiu, A., Jastrzebski, S., Morrone, B., De Laroussilhe, Q., Gesmundo, A., Attariyan, M., Gelly, S.: Parameter-efficient transfer learning for NLP. In: International Conference on Machine Learning (ICML), pp. 2790–2799 (2019)
8. Ben Zaken, E., Goldberg, Y., Ravfogel, S.: BitFit: simple parameter-efficient fine-tuning for transformer-based masked language-models. In: Annual Meeting of the Association for Computational Linguistics (ACL), pp. 1–9 (2022)
9. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., et al.: An image is worth 16x16 words: transformers for image recognition at scale. In: International Conference on Learning Representations (ICLR) (2021)
10. Huang, Y., Qiu, C., Yuan, K.: Surface defect saliency of magnetic tile. Vis. Comput. 36, 85–96 https://doi.org/10.1007/s00371-018-1588-5. Dataset: (2020). https://github.com/abin24/Magnetic-tile-defect-datasets
11. NEU-Seg dataset: GitHub repository. (2026). https://github.com/DHW-Master/NEU_Seg (accessed 27
12. Liu, Y., Zhang, C., Dong, X.: A survey of real-time surface defect inspection methods based on deep learning. Artif. Intell. Rev. **56**, 12131–12170 (2023). https://doi.org/10.1007/s10462-023-10475-7
13. Li, D., Niu, Z., Peng, D.: Magnetic tile surface defect detection based on texture feature clustering. J. Shanghai Jiaotong Univ. (Sci). **24**, 663–670 (2019). https://doi.org/10.1007/s12204-019-2101-7
14. Xie, L., Lin, L., Yin, M., Meng, L., Yin, G.: A novel surface defect inspection algorithm for magnetic tile. Appl. Surf. Sci. **375**, 118–126 (2016). https://doi.org/10.1016/j.apsusc.2016.03.013
15. Liang, W., Sun, Y.: ELCNN: a deep neural network for small object defect detection of magnetic tile. IEEE Trans. Instrum. Meas. **71**, 1–10 (2022). https://doi.org/10.1109/TIM.2022.3193175
16. Zhong, Z., Wang, H., Xiang, D.: Small defect detection based on local structure similarity for magnetic tile surface. Electronics. **12**, 185 (2022). https://doi.org/10.3390/electronics12010185

17. Neven, R., Goedemé, T.: A multi-branch U-Net for steel surface defect type and severity segmentation. Metals. **11**, 870 (2021). https://doi.org/10.3390/met11060870
18. Li, C., Yan, H., Qian, X., Zhu, S., Zhu, P., Liao, C., Tian, H., Li, X., Wang, X., Li, X.: A domain adaptation YOLOv5 model for industrial defect inspection. Measurement. **213**, 112725 (2023). https://doi.org/10.1016/j.measurement.2023.112725
19. Jia, M., Tang, L., Chen, B.C., Cardie, C., Belongie, S., Hariharan, B., Lim, S.N.: Visual prompt tuning. In: European Conference on Computer Vision (ECCV), pp. 709–727 (2022). https://doi.org/10.1007/978-3-031-19827-4_41
20. Chen, S., Ge, C., Tong, Z., Wang, J., Song, Y., Wang, J., Luo, P.: AdaptFormer: adapting vision transformers for scalable visual recognition. In: Advances in Neural Information Processing Systems (NeurIPS), pp. 16664–16678 (2022)
21. Jie, S., Deng, Z.H., Chen, S., Jin, Z.: Convolutional bypasses are better vision transformer adapters. In: Frontiers in Artificial Intelligence and Applications (2024). https://doi.org/10.3233/FAIA240489
22. He, J., Zhou, C., Ma, X., Berg-Kirkpatrick, T., Neubig, G.: Towards a unified view of parameter-efficient transfer learning. In: International Conference on Learning Representations (ICLR) arXiv:2110.04366 (2022)
23. De Marinis, P., Vessio, G., Castellano, G.: Take a peek: efficient encoder adaptation for few-shot semantic segmentation via LoRA. Pattern Recognit. Lett. **207**, 47–54 (2026). https://doi.org/10.1016/j.patrec.2026.06.003
24. Lialin, V., Deshpande, V., Rumshisky, A.: Scaling down to scale up: a guide to parameter-efficient fine-tuning arXiv:2303.15647 (2023)
25. Geirhos, R., Jacobsen, J.H., Michaelis, C., Zemel, R., Brendel, W., Bethge, M., Wichmann, F.A.: Shortcut learning in deep neural networks. Nat. Mach. Intell. **2**, 665–673 (2020). https://doi.org/10.1038/s42256-020-00257-z
26. Kapoor, S., Narayanan, A.: Leakage and the reproducibility crisis in machine-learning-based science. Patterns. **4**, 100804 (2023). https://doi.org/10.1016/j.patter.2023.100804
27. Zhou, K., Liu, Z., Qiao, Y., Xiang, T., Loy, C.C.: Domain generalization: a survey. IEEE Trans. Pattern Anal. Mach. Intell. **45**, 4396–4415 (2023). https://doi.org/10.1109/TPAMI.2022.3195549
28. Sagawa, S., Koh, P.W., Hashimoto, T.B., Liang, P.: Distributionally robust neural networks for group shifts: on the importance of regularization for worst-case generalization. In: International Conference on Learning Representations (ICLR) arXiv:1911.08731 (2020)
29. Hossain, M.S., Betts, J.M., Paplinski, A.P.: Dual focal loss to address class imbalance in semantic segmentation. Neurocomputing. **462**, 69–87 (2021). https://doi.org/10.1016/j.neucom.2021.07.055
30. Sudre, C.H., Li, W., Vercauteren, T., Ourselin, S., Cardoso, M.J.: Generalised Dice overlap as a deep learning loss function for highly unbalanced segmentations. In: MICCAI Workshop on Machine Learning in Medical Imaging, LNCS 10553, pp. 240–248 (2017). https://doi.org/10.1007/978-3-319-67558-9_28

31. Lin, T.Y., Goyal, P., Girshick, R., He, K., Dollár, P.: Focal loss for dense object detection. In: IEEE International Conference on Computer Vision (ICCV), pp. 2999–3007 (2017). https://doi.org/10.1109/ICCV.2017.324
32. Milletari, F., Navab, N., Ahmadi, S.A.: V-Net: fully convolutional neural networks for volumetric medical image segmentation. In: Fourth International Conference on 3D Vision (3DV), pp. 565–571 (2016). https://doi.org/10.1109/3DV.2016.79
33. Pfeiffer, J., Kamath, A., Rücklé, A., Cho, K., Gurevych, I.: AdapterFusion: non-destructive task composition for transfer learning. In: Conference of the European Chapter of the Association for Computational Linguistics (EACL), pp. 487–503 (2021). https://doi.org/10.18653/v1/2021.eacl-main.39
34. Karimi Mahabadi, R., Henderson, J., Ruder, S.: Compacter: efficient low-rank hypercomplex adapter layers. In: Advances in Neural Information Processing Systems (NeurIPS) arXiv:2106.04647 (2021)
35. Buda, M., Maki, A., Mazurowski, M.A.: A systematic study of the class imbalance problem in convolutional neural networks. Neural Netw. **106**, 249–259 (2018). https://doi.org/10.1016/j.neunet.2018.07.011

Figures

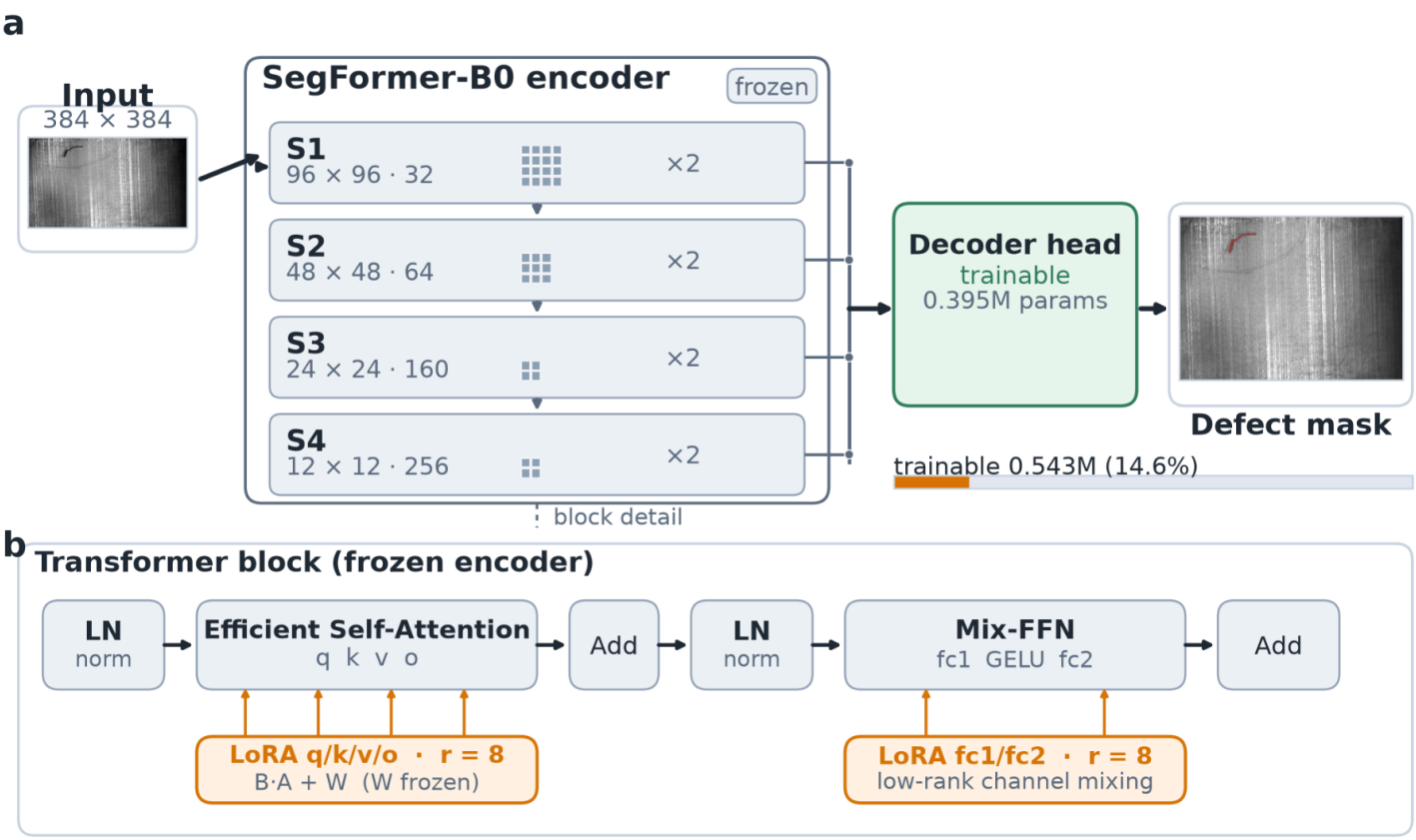


Figure 1

LoRA-adapted SegFormer-B0. The four encoder stages feed a multi-scale feature bus to the decoder; each stage contains two transformer blocks. The encoder stays frozen, and low-rank matrices in q/k/v/o and the MLP projections, plus the decoder head, are trained.

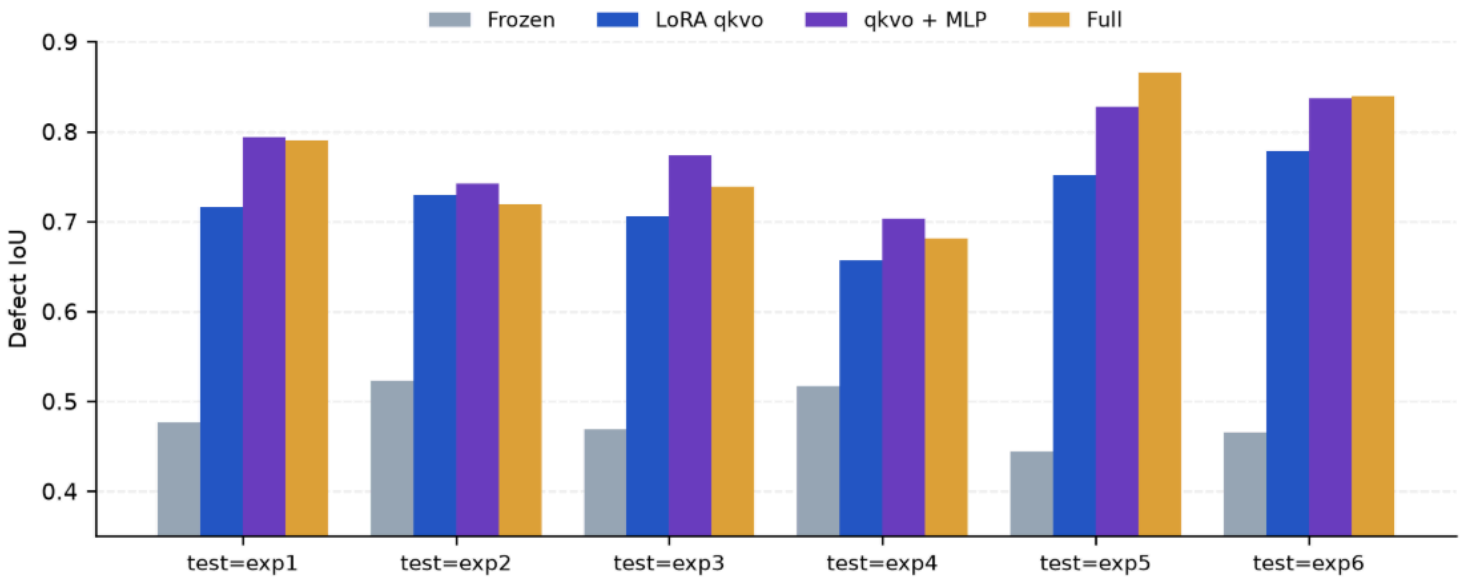


Figure 2

Per-group defect IoU on MTD (3-seed means). Adding MLP adapters raises the bars above attention-only LoRA and, on four of six groups, above Full. Seed-level means are reported in Appendix A.

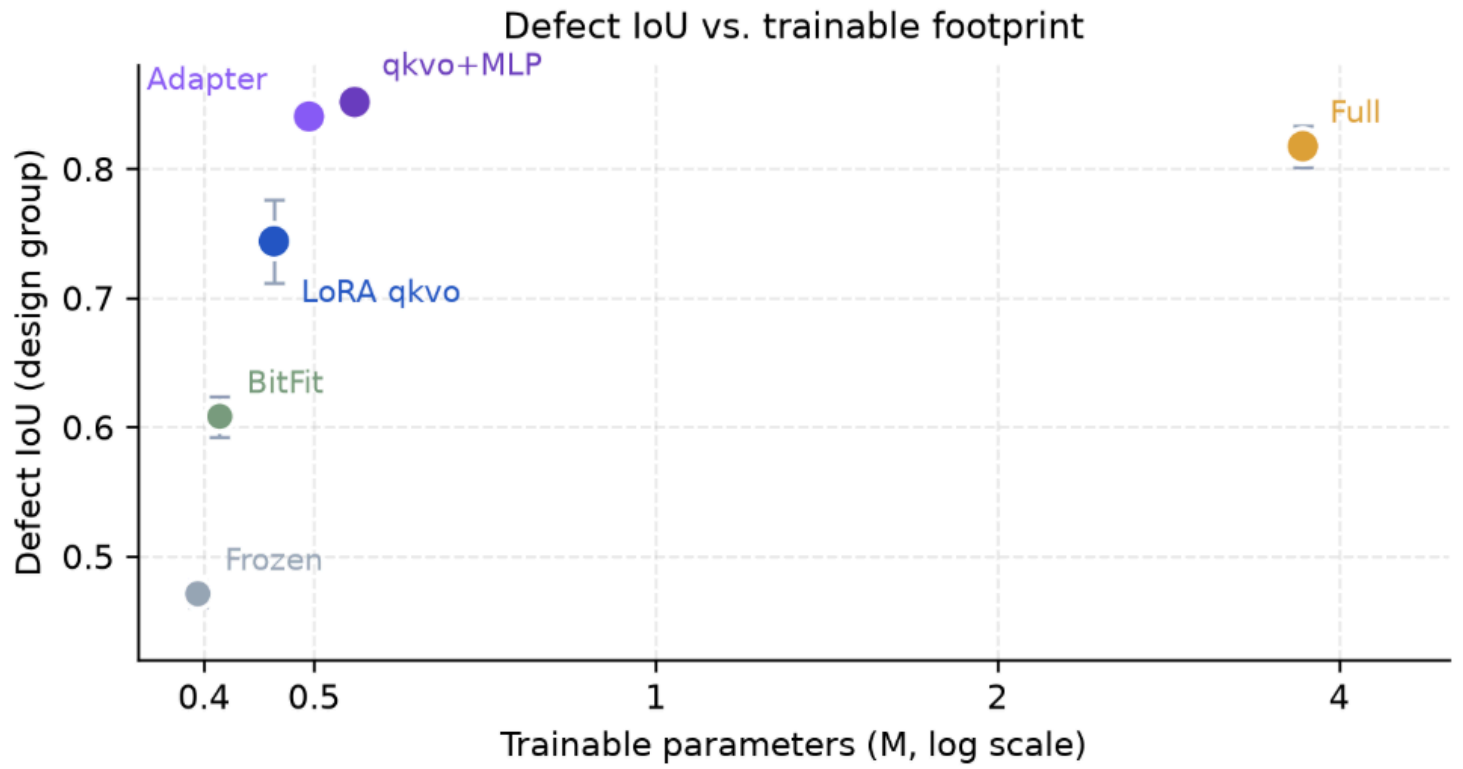


Figure 3

Accuracy vs trainable parameters on the MTD exp6 design-selection group. qkvo+MLP sits above Full with 0.54M parameters; attention-only LoRA is below Full; the adapter is above Full with overlapping uncertainty. DeepLabV3 is excluded because it did not converge under the shared budget.

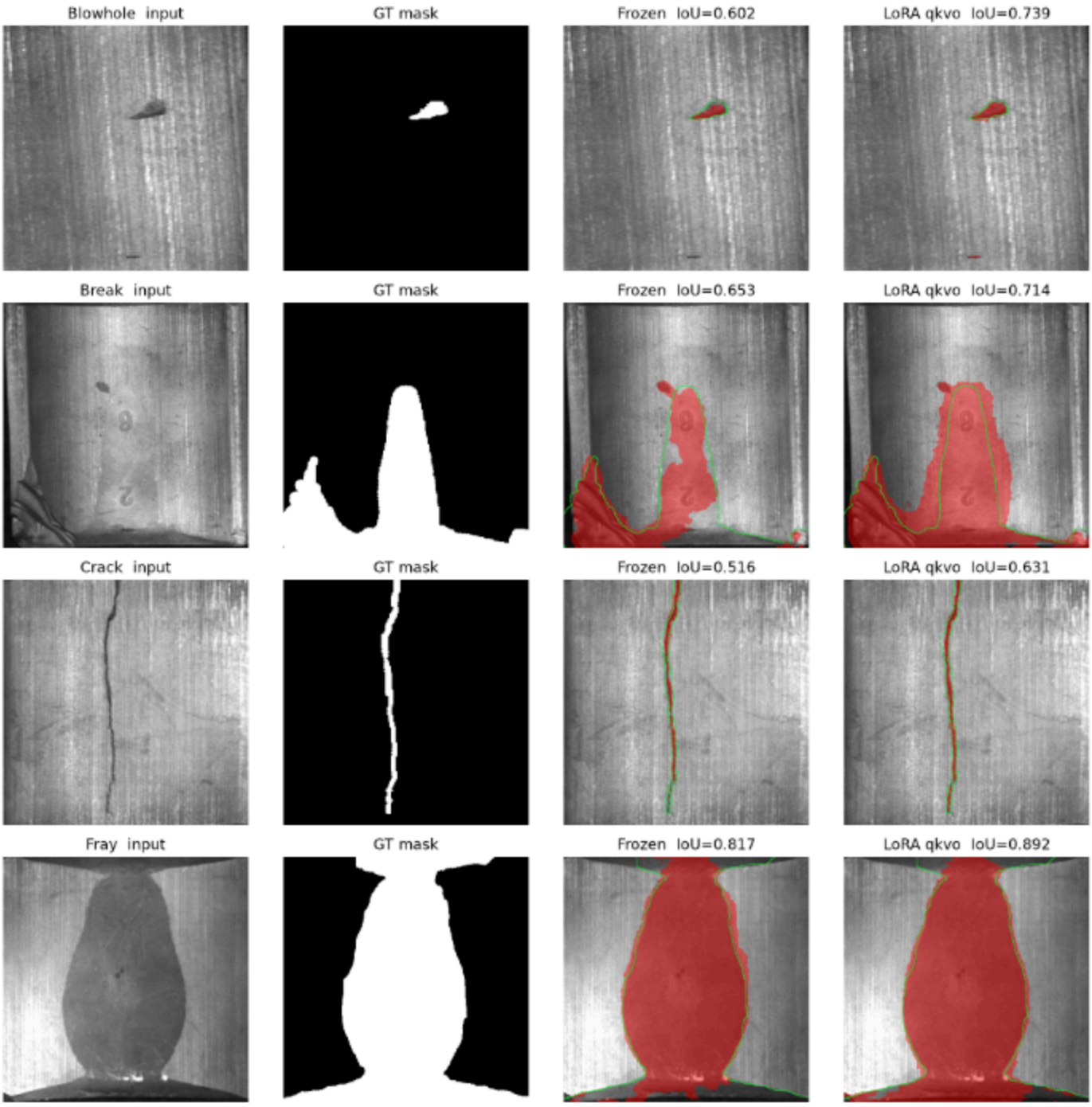


Figure 4

Qualitative comparison on the exp6 design-selection group (attention-only LoRA vs Frozen). Green contours mark ground truth; red regions show predictions. Rows: blowhole, break, crack, fray.

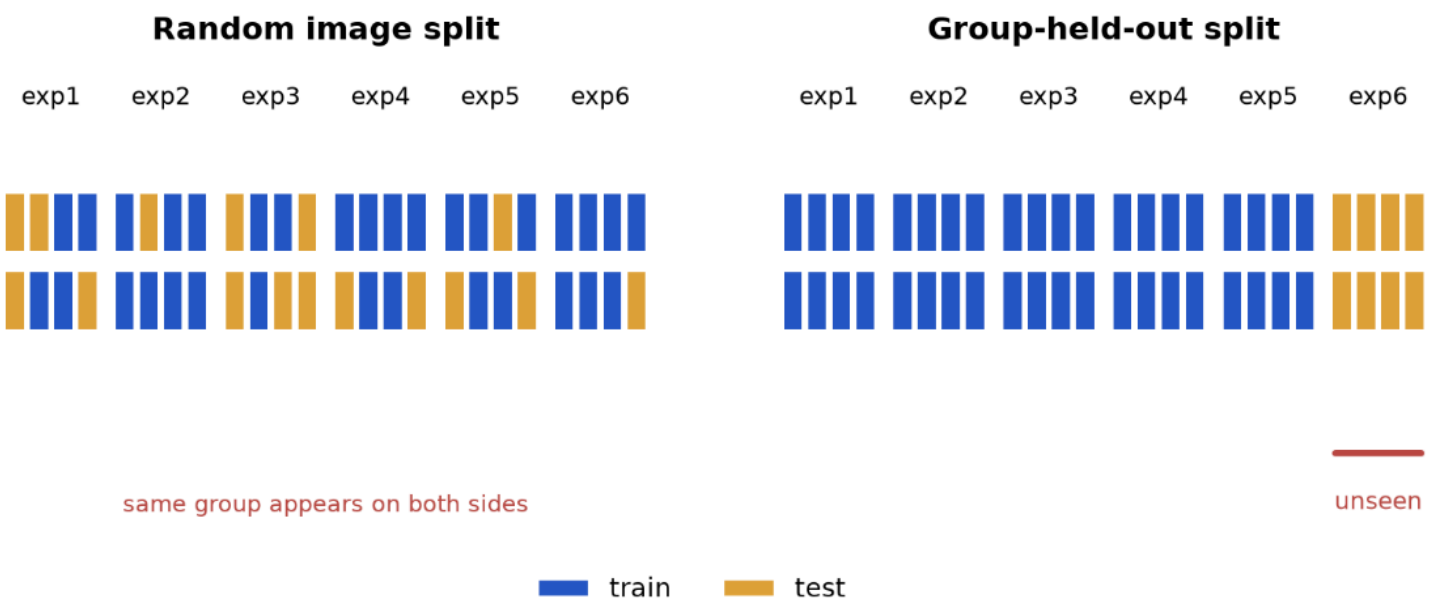


Figure 5

Schematic of the two split constructions. Random image splits may place related crops on both sides; group-held-out splits keep one whole filename group unseen. The mixing is a plausible contributor, not a measured mechanism.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Appendixs.docx