# Supplementary Information

Comparing spectral-robust training objectives in physics-consistent dual-wavelength diffractive networks

Zelin Lu[1]*
[1]China Mobile Research Institute, Beijing 100053, China
*Correspondence: luzelin@chinamobile.com

This file reports supplementary evidence obtained with the same physical thickness model, training configurations, and statistical definitions as the main text. All run-level values underlying the figures and tables are provided in Source_Data.xlsx.

## S1. Supplementary methods and reproducibility

The manuscript and Source_Data.xlsx use the same frozen data version. One complete independent training run is the statistical unit, and the two primary objectives are paired by seed. The broad benchmark contains eight configurations with ten runs each; direct 128-grid training contains ten runs per objective; the physical-domain study contains 62 runs.

**Supplementary Table S1. Architecture, readout, and training objective of the eight configurations.**

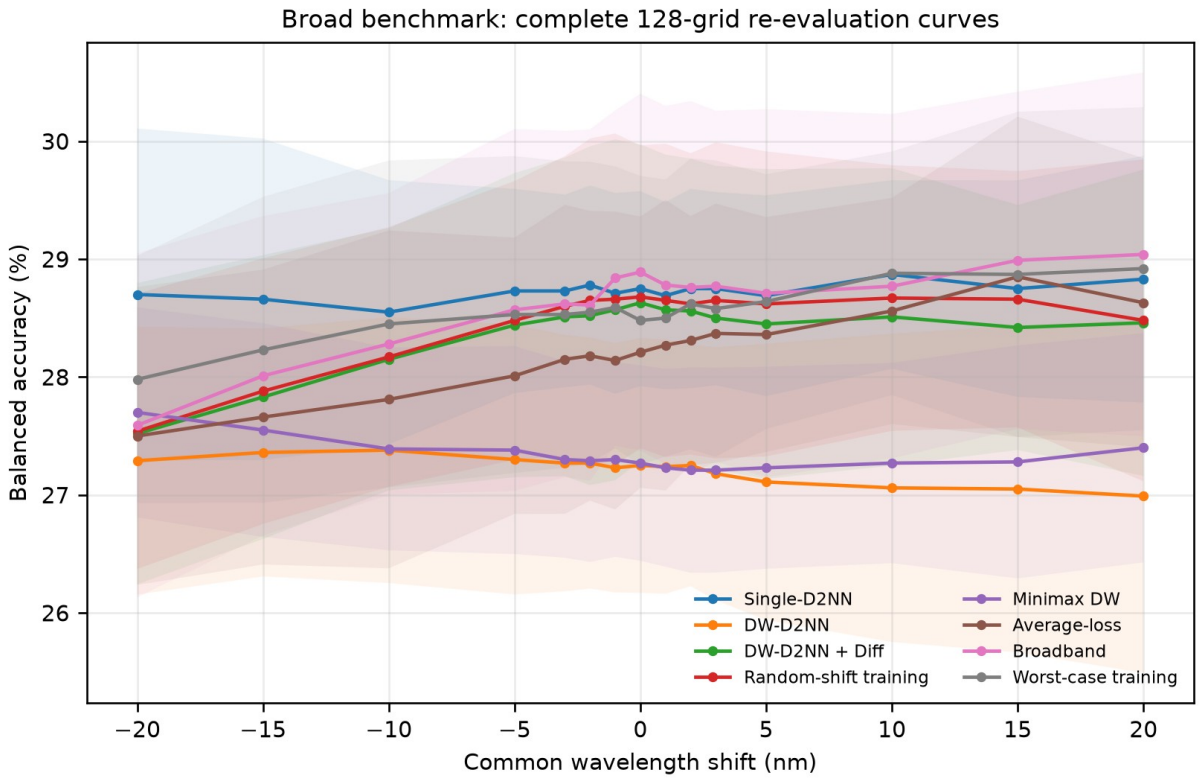
| Display name | Internal ID | Wavelength channels | Differential readout | Training objective |
|---|---|---|---|---|
| Single-D2NN | single | 532 nm | no | Nominal-condition cross-entropy |
| DW-D2NN | dw | 532/633 nm | no; mean of two intensities | Nominal-condition cross-entropy |
| DW-D2NN + Diff | dw_diff | 532/633 nm | yes | Nominal-condition cross-entropy |
| Randomized wavelength-perturbation training | nadw_random | 532/633 nm | yes | Nominal loss plus Gaussian wavelength-perturbation loss |
| Average multi-shift training | average_shift | 532/633 nm | yes | Mean cross-entropy over seven fixed offsets |
| Spectral-mean training | broadband | 532/633 nm | yes | Average logits over seven offsets, then compute cross-entropy |
| Sample-wise worst-case training | nadw_worst | 532/633 nm | yes | Maximum per-sample cross-entropy over seven fixed offsets |
| Non-differential minimax training | minimax_dw | 532/633 nm | no; mean of two intensities | Maximum per-sample cross-entropy over seven fixed offsets |

**Supplementary Table S2. GPU environment and deterministic settings.**

| key | value |
|---|---|

| | |
|---|---|
| python | 3.11.9 (tags/v3.11.9:de54cf5, Apr 2 2024, 10:12:12) [MSC v.1938 64 bit (AMD64)] |
| platform | Windows-10-10.0.26200-SP0 |
| torch | 2.12.0+cu130 |
| torchvision | 0.27.0+cu130 |
| cuda_runtime | 13.0 |
| cuda_available | True |
| gpu | NVIDIA GeForce RTX 5080 Laptop GPU |
| gpu_capability | [12, 0] |
| nvidia_smi | NVIDIA GeForce RTX 5080 Laptop GPU, 592.01, 16303 MiB |
| numpy | 2.4.4 |
| scipy | 1.17.1 |
| scikit_learn | 1.9.0 |
| medmnist | 3.0.1 |
| deterministic_algorithms_used | True |
| mixed_precision_used | False |

## S2. Broad benchmark and paired statistics




**Supplementary Figure S1. Complete wavelength-drift curves for eight configurations trained on the 64-grid and re-evaluated on the 128-grid. Lines show means over ten independent training runs; shading indicates s.d.**

**Supplementary Table S3. Method-level summary of the broad benchmark.**

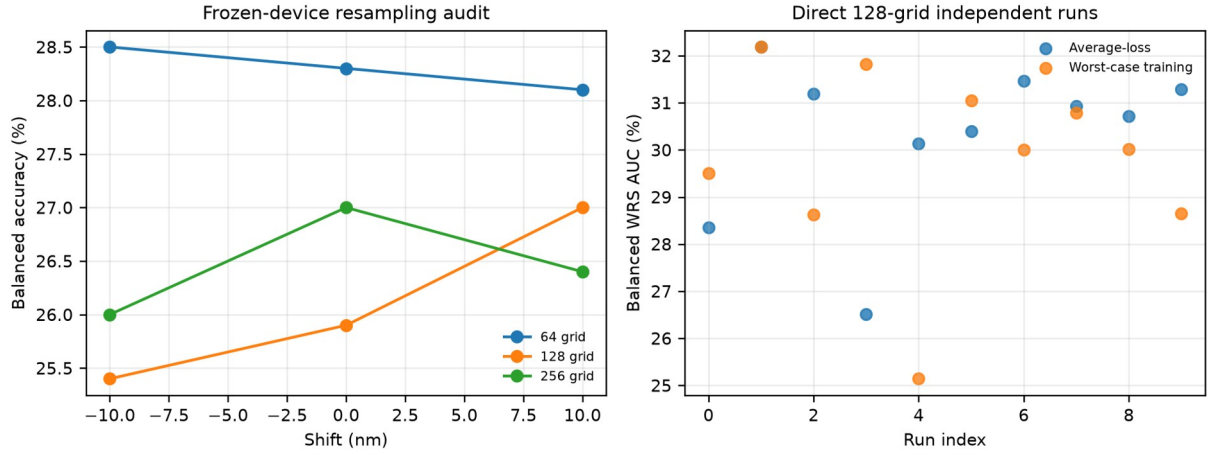
| method | grid_size | clean_balanced_accuracy_mean | balanced_wrs_auc_mean | balanced_wrs_auc_std | worst_shift_accuracy_mean |
|---|---|---|---|---|---|
| average_shift | 128 | 0.2821 | 0.281977 | 0.00990153 | 0.272 |
| broadband | 128 | 0.2889 | 0.285567 | 0.0131347 | 0.2742 |

| | | | | | |
|---|---|---|---|---|---|
| dw | 128 | 0.2725 | 0.272075 | 0.0114429 | 0.2656 |
| dw_diff | 128 | 0.2863 | 0.282995 | 0.010682 | 0.2702 |
| minimax_dw | 128 | 0.2727 | 0.2736 | 0.00824077 | 0.2692 |
| nadw_random | 128 | 0.2868 | 0.283984 | 0.00987021 | 0.2704 |
| nadw_worst | 128 | 0.2848 | 0.285725 | 0.0117948 | 0.2771 |
| single | 128 | 0.2875 | 0.287211 | 0.00939465 | 0.2814 |

**Supplementary Table S4. Complete paired statistics for the broad benchmark.**

| comparison | metric | n_independent_training_runs | mean_paired_difference | bootstrap_ci95_lower | bootstrap_ci95_upper | exact_sign_flip_pvalue_two_sided |
|---|---|---|---|---|---|---|
| nadw_worst - average_shift | wrs_auc | 10 | 0.0037475 | -0.00549628 | 0.0127826 | 0.464844 |
| nadw_worst - average_shift | balanced_wrs_auc | 10 | 0.0037475 | -0.0054575 | 0.01279 | 0.464844 |
| nadw_worst - average_shift | worst_shift_accuracy | 10 | 0.0051 | -0.005 | 0.0155 | 0.402344 |
| nadw_worst - broadband | wrs_auc | 10 | 0.0001575 | -0.01265 | 0.01215 | 0.982422 |
| nadw_worst - broadband | balanced_wrs_auc | 10 | 0.0001575 | -0.0126275 | 0.0122125 | 0.982422 |
| nadw_worst - broadband | worst_shift_accuracy | 10 | 0.0029 | -0.0115 | 0.0168 | 0.722656 |
| nadw_worst - dw | wrs_auc | 10 | 0.01365 | 0.0048075 | 0.0221162 | 0.0234375 |
| nadw_worst - dw | balanced_wrs_auc | 10 | 0.01365 | 0.00472375 | 0.0220837 | 0.0234375 |
| nadw_worst - dw | worst_shift_accuracy | 10 | 0.0115 | 0.003 | 0.0198 | 0.0390625 |
| nadw_worst - dw_diff | wrs_auc | 10 | 0.00273 | -0.0080925 | 0.0132975 | 0.636719 |
| nadw_worst - dw_diff | balanced_wrs_auc | 10 | 0.00273 | -0.00812884 | 0.0132525 | 0.636719 |
| nadw_worst - dw_diff | worst_shift_accuracy | 10 | 0.0069 | -0.0031 | 0.0176 | 0.259766 |
| nadw_worst - minimax_dw | wrs_auc | 10 | 0.012125 | 0.00536375 | 0.01835 | 0.0136719 |
| nadw_worst - minimax_dw | balanced_wrs_auc | 10 | 0.012125 | 0.00536875 | 0.01837 | 0.0136719 |
| nadw_worst - minimax_dw | worst_shift_accuracy | 10 | 0.0079 | 0.0009 | 0.015 | 0.0742188 |
| nadw_worst - nadw_random | wrs_auc | 10 | 0.00174125 | -0.0087925 | 0.0117088 | 0.751953 |
| nadw_worst - nadw_random | balanced_wrs_auc | 10 | 0.00174125 | -0.00868 | 0.0117138 | 0.751953 |
| nadw_worst - nadw_random | worst_shift_accuracy | 10 | 0.0067 | -0.0026 | 0.0166 | 0.236328 |
| nadw_worst - single | wrs_auc | 10 | -0.00148625 | -0.010575 | 0.00535378 | 0.849609 |
| nadw_worst - single | balanced_wrs_auc | 10 | -0.00148625 | -0.0105212 | 0.00535 | 0.849609 |
| nadw_worst - single | worst_shift_accuracy | 10 | -0.0043 | -0.0137 | 0.0034 | 0.455078 |

## S3. Numerical convergence and direct optimization




**Supplementary Figure S2. Difference between frozen-device resampling audits and direct 128-grid training. The left panel changes only propagation sampling; the right panel reports direct optimization from random initialization.**

**Supplementary Table S5. Frozen-device resampling on 64-, 128-, and 256-grid propagation meshes.**

| grid_size | shift_nm | accuracy | balanced_accuracy | macro_f1 | macro_auc | minimum_class_recall |
|---|---|---|---|---|---|---|
| 64 | -10 | 0.285 | 0.285 | 0.221264 | 0.600403 | 0.016 |
| 64 | 0 | 0.283 | 0.283 | 0.218106 | 0.601601 | 0.02 |
| 64 | 10 | 0.281 | 0.281 | 0.217787 | 0.600035 | 0.02 |
| 128 | -10 | 0.254 | 0.254 | 0.200533 | 0.584835 | 0.004 |
| 128 | 0 | 0.259 | 0.259 | 0.205158 | 0.58831 | 0.008 |
| 128 | 10 | 0.27 | 0.27 | 0.218832 | 0.586285 | 0.008 |
| 256 | -10 | 0.26 | 0.26 | 0.209041 | 0.581759 | 0.008 |
| 256 | 0 | 0.27 | 0.27 | 0.21903 | 0.581887 | 0.008 |
| 256 | 10 | 0.264 | 0.264 | 0.213577 | 0.582162 | 0.004 |

**Supplementary Table S6. Zero-padding audit.**

| padding_factor | prediction_agreement_with_unpadded | relative_logit_l2 |
|---|---|---|
| 1 | 1 | 0 |
| 2 | 1 | 0.00720732 |
| 4 | 1 | 0.00793403 |

**Supplementary Table S7. Floating-point precision audit.**

| comparison | prediction_agreement | relative_logit_l2 |
|---|---|---|
| float32_vs_float64 | 1 | 0.000363959 |

**Supplementary Table S8. Layer-wise energy and finite-aperture transmission.**

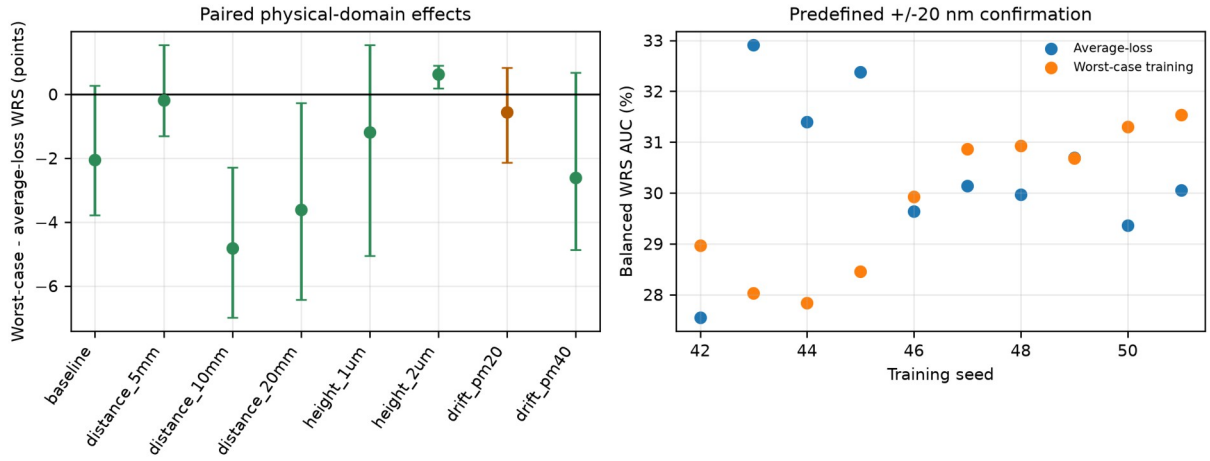
| wavelength_nm | stage | energy | aperture_transmission_ratio | propagation_energy_ratio |
|---|---|---|---|---|
| 532 | input | 5067.87 | nan | nan |
| 532 | layer_1 | 1895.69 | 0.374059 | 1 |
| 532 | layer_2 | 1716.17 | 0.905302 | 1 |
| 532 | layer_3 | 1593.98 | 0.9288 | 1 |
| 633 | input | 5067.87 | nan | nan |
| 633 | layer_1 | 1895.69 | 0.37406 | 1 |
| 633 | layer_2 | 1711.57 | 0.902878 | 1 |
| 633 | layer_3 | 1562.61 | 0.91297 | 1 |

**Supplementary Table S9. All 20 direct 128-grid training runs.**

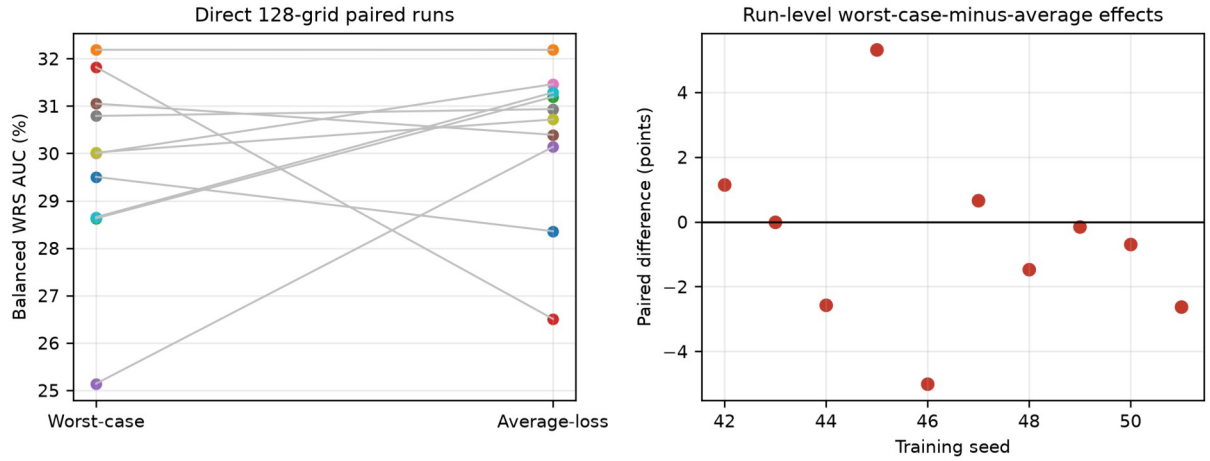
| method | seed | best_epoch | clean_balanced_accura | balanced_wrs_auc | worst_shift_accuracy |
|---|---|---|---|---|---|

| | | | cy | | |
|---|---|---|---|---|---|
| nadw_worst | 42 | 3 | 0.298 | 0.29505 | 0.284 |
| nadw_worst | 43 | 19 | 0.325 | 0.321862 | 0.311 |
| nadw_worst | 44 | 4 | 0.294 | 0.286287 | 0.278 |
| nadw_worst | 45 | 14 | 0.321 | 0.318213 | 0.31 |
| nadw_worst | 46 | 4 | 0.25 | 0.2514 | 0.249 |
| nadw_worst | 47 | 20 | 0.316 | 0.310513 | 0.3 |
| nadw_worst | 48 | 4 | 0.306 | 0.3 | 0.282 |
| nadw_worst | 49 | 8 | 0.31 | 0.3079 | 0.305 |
| nadw_worst | 50 | 20 | 0.301 | 0.300212 | 0.297 |
| nadw_worst | 51 | 7 | 0.291 | 0.28655 | 0.277 |
| average_shift | 42 | 3 | 0.29 | 0.283575 | 0.274 |
| average_shift | 43 | 20 | 0.325 | 0.32185 | 0.314 |
| average_shift | 44 | 17 | 0.319 | 0.311937 | 0.297 |
| average_shift | 45 | 4 | 0.261 | 0.265062 | 0.26 |
| average_shift | 46 | 19 | 0.302 | 0.301375 | 0.292 |
| average_shift | 47 | 5 | 0.304 | 0.303925 | 0.295 |
| average_shift | 48 | 13 | 0.315 | 0.314612 | 0.299 |
| average_shift | 49 | 19 | 0.304 | 0.309337 | 0.293 |
| average_shift | 50 | 20 | 0.309 | 0.307163 | 0.293 |
| average_shift | 51 | 10 | 0.317 | 0.31285 | 0.296 |

## S4. Prespecified physical-domain analysis




**Supplementary Figure S3. Paired effects in the prespecified physical domain and the ten-seed confirmation for the ±20 nm training range. Error bars are paired-seed bootstrap 95% confidence intervals.**




**Supplementary Figure S4. Seed-level direct 128-grid results and raw sample-wise-worst-case-minus-average-loss differences.**

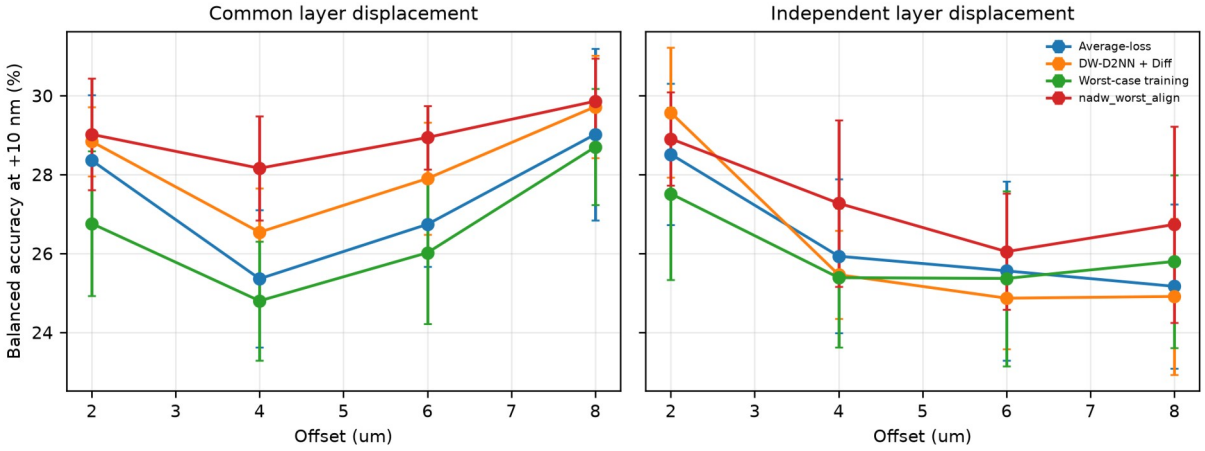
**Supplementary Table S10. Paired Balanced-WRS AUC statistics in the physical domain.**

| experiment | role | metric | n_paired_seeds | mean_paired_difference_nadw_minus_average | bootstrap_ci95_lower | bootstrap_ci95_upper | exact_sign_flip_pvalue_two_sided |
|---|---|---|---|---|---|---|---|
| baseline | three_seed_screen | balanced_wrs_auc | 3 | -0.0204687 | -0.0377812 | 0.00278125 | 0.5 |
| distance_5mm | three_seed_screen | balanced_wrs_auc | 3 | -0.00179167 | -0.013125 | 0.0154687 | 1 |
| distance_10mm | three_seed_screen | balanced_wrs_auc | 3 | -0.0481771 | -0.0699375 | -0.0228437 | 0.25 |
| distance_20mm | three_seed_screen | balanced_wrs_auc | 3 | -0.0361667 | -0.0643125 | -0.0026875 | 0.25 |
| height_1um | three_seed_screen | balanced_wrs_auc | 3 | -0.0119583 | -0.0505625 | 0.015375 | 0.75 |
| height_2um | three_seed_screen | balanced_wrs_auc | 3 | 0.0063125 | 0.00184375 | 0.00896875 | 0.25 |
| drift_pm20 | predefined_confirmation | balanced_wrs_auc | 10 | -0.00556875 | -0.0213063 | 0.008325 | 0.498047 |
| drift_pm40 | three_seed_screen | balanced_wrs_auc | 3 | -0.0261042 | -0.0487187 | 0.00684375 | 0.5 |

**Supplementary Table S11. Mean and s.d. by configuration in the physical domain.**

| id | role | method | count | mean | std |
|---|---|---|---|---|---|
| baseline | three_seed_screen | average_shift | 3 | 0.30324 | 0.0274195 |
| baseline | three_seed_screen | nadw_worst | 3 | 0.282771 | 0.0152168 |
| distance_10mm | three_seed_screen | average_shift | 3 | 0.398083 | 0.0196737 |
| distance_10mm | three_seed_screen | nadw_worst | 3 | 0.349906 | 0.0434029 |
| distance_20mm | three_seed_screen | average_shift | 3 | 0.372792 | 0.0246846 |
| distance_20mm | three_seed_screen | nadw_worst | 3 | 0.336625 | 0.0160414 |
| distance_5mm | three_seed_screen | average_shift | 3 | 0.392646 | 0.00510192 |
| distance_5mm | three_seed_screen | nadw_worst | 3 | 0.390854 | 0.0120139 |
| drift_pm20 | predefined_confirmation | average_shift | 10 | 0.304091 | 0.0154497 |
| drift_pm20 | predefined_confirmation | nadw_worst | 10 | 0.298522 | 0.014121 |
| drift_pm40 | three_seed_screen | average_shift | 3 | 0.307615 | 0.0230951 |
| drift_pm40 | three_seed_screen | nadw_worst | 3 | 0.28151 | 0.00781877 |
| height_1um | three_seed_screen | average_shift | 3 | 0.292083 | 0.0296039 |
| height_1um | three_seed_screen | nadw_worst | 3 | 0.280125 | 0.0229104 |
| height_2um | three_seed_screen | average_shift | 3 | 0.312104 | 0.00416368 |
| height_2um | three_seed_screen | nadw_worst | 3 | 0.318417 | 0.00664078 |

## S5. Alignment failure boundary



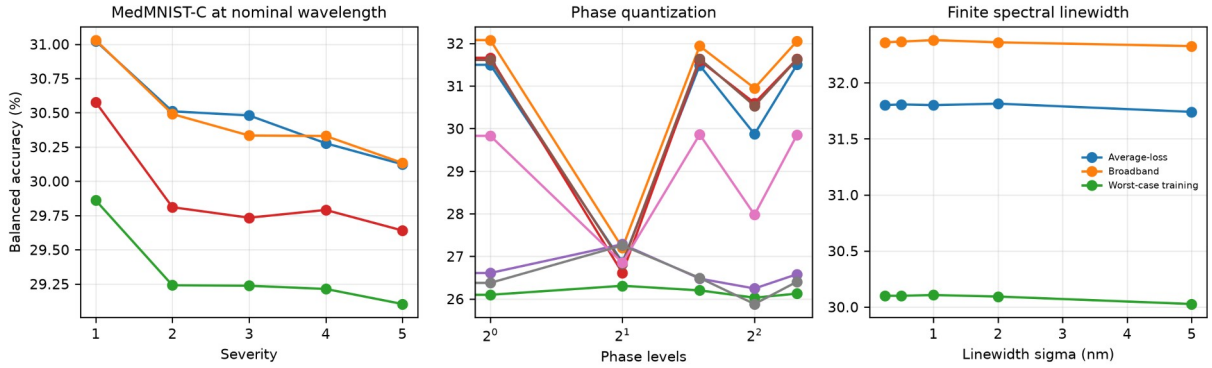

**Supplementary Figure S5. Complete common-layer and independent-layer displacement curves at +10 nm wavelength drift. Points show means over independent training runs and error bars show s.d.**

**Supplementary Table S12. Displacement summary at +10 nm.**

| method | alignment_type | offset_um | accuracy_mean | accuracy_std | balanced_accuracy_mean | balanced_accuracy_std |
|---|---|---|---|---|---|---|
| average_shift | common | 2 | 0.2836 | 0.0165318 | 0.2836 | 0.0165318 |
| average_shift | common | 4 | 0.2536 | 0.0174155 | 0.2536 | 0.0174155 |
| average_shift | common | 6 | 0.2674 | 0.0107145 | 0.2674 | 0.0107145 |
| average_shift | common | 8 | 0.2902 | 0.0217532 | 0.2902 | 0.0217532 |
| average_shift | independent | 2 | 0.28516 | 0.0178761 | 0.28516 | 0.0178761 |
| average_shift | independent | 4 | 0.25932 | 0.0195122 | 0.25932 | 0.0195122 |
| average_shift | independent | 6 | 0.2556 | 0.0226752 | 0.2556 | 0.0226752 |
| average_shift | independent | 8 | 0.25168 | 0.0207961 | 0.25168 | 0.0207961 |
| average_shift | registered | 0 | 0.3172 | 0.0222306 | 0.3172 | 0.0222306 |
| dw_diff | common | 2 | 0.2884 | 0.00876356 | 0.2884 | 0.00876356 |
| dw_diff | common | 4 | 0.2654 | 0.0111041 | 0.2654 | 0.0111041 |
| dw_diff | common | 6 | 0.279 | 0.0142127 | 0.279 | 0.0142127 |
| dw_diff | common | 8 | 0.2972 | 0.0129692 | 0.2972 | 0.0129692 |
| dw_diff | independent | 2 | 0.29572 | 0.0164609 | 0.29572 | 0.0164609 |
| dw_diff | independent | 4 | 0.25464 | 0.0111388 | 0.25464 | 0.0111388 |
| dw_diff | independent | 6 | 0.24868 | 0.0128961 | 0.24868 | 0.0128961 |
| dw_diff | independent | 8 | 0.24912 | 0.0197786 | 0.24912 | 0.0197786 |
| dw_diff | registered | 0 | 0.3196 | 0.00944987 | 0.3196 | 0.00944987 |
| nadw_worst | common | 2 | 0.2676 | 0.0183112 | 0.2676 | 0.0183112 |
| nadw_worst | common | 4 | 0.248 | 0.0150499 | 0.248 | 0.0150499 |
| nadw_worst | common | 6 | 0.2602 | 0.0180194 | 0.2602 | 0.0180194 |
| nadw_worst | common | 8 | 0.287 | 0.0146969 | 0.287 | 0.0146969 |
| nadw_worst | independent | 2 | 0.27516 | 0.0218187 | 0.27516 | 0.0218187 |
| nadw_worst | independent | 4 | 0.25388 | 0.0176454 | 0.25388 | 0.0176454 |
| nadw_worst | independent | 6 | 0.25368 | 0.0221712 | 0.25368 | 0.0221712 |
| nadw_worst | independent | 8 | 0.258 | 0.0218594 | 0.258 | 0.0218594 |
| nadw_worst | registered | 0 | 0.2992 | 0.0195371 | 0.2992 | 0.0195371 |
| nadw_worst_align | common | 2 | 0.2902 | 0.0141315 | 0.2902 | 0.0141315 |
| nadw_worst_align | common | 4 | 0.2816 | 0.0131833 | 0.2816 | 0.0131833 |
| nadw_worst_align | common | 6 | 0.2894 | 0.00804984 | 0.2894 | 0.00804984 |

| | | | | | | |
|---|---|---|---|---|---|---|
| nadw_worst_align | common | 8 | 0.2986 | 0.0108536 | 0.2986 | 0.0108536 |
| nadw_worst_align | independent | 2 | 0.28908 | 0.0118495 | 0.28908 | 0.0118495 |
| nadw_worst_align | independent | 4 | 0.27272 | 0.0211138 | 0.27272 | 0.0211138 |
| nadw_worst_align | independent | 6 | 0.26048 | 0.0147058 | 0.26048 | 0.0147058 |
| nadw_worst_align | independent | 8 | 0.26736 | 0.0248728 | 0.26736 | 0.0248728 |
| nadw_worst_align | registered | 0 | 0.3126 | 0.00512835 | 0.3126 | 0.00512835 |

## S6. Input corruption, phase quantization, and linewidth




**Supplementary Figure S6. Supplementary failure boundaries under MedMNIST-C corruption, phase quantization, and finite spectral linewidth. Panels aggregate over the corresponding training runs, wavelength shifts, or corruption types; complete run-level data are in Source Data.**

The MedMNIST-C, Phase_quantization, and Finite_linewidth sheets contain 3,000, 1,200, and 225 run-condition records, respectively. This section reports prespecified summaries and does not extrapolate input-corruption results to general hardware robustness.

## S7. Compact electronic reference

**Supplementary Table S13. Compact CNN results across ten independent training runs.**

| model | seed | parameter_count | best_epoch | accuracy | balanced_accuracy | macro_f1 | macro_auc |
|---|---|---|---|---|---|---|---|
| CompactCNN end-to-end | 42 | 4924 | 27 | 0.779 | 0.779 | 0.78049 | 0.969185 |
| CompactCNN end-to-end | 43 | 4924 | 3 | 0.684 | 0.684 | 0.675192 | 0.930104 |
| CompactCNN end-to-end | 44 | 4924 | 15 | 0.774 | 0.774 | 0.775306 | 0.971184 |
| CompactCNN end-to-end | 45 | 4924 | 9 | 0.752 | 0.752 | 0.751983 | 0.956151 |
| CompactCNN end-to-end | 46 | 4924 | 5 | 0.703 | 0.703 | 0.701546 | 0.951707 |
| CompactCNN end-to-end | 47 | 4924 | 17 | 0.761 | 0.761 | 0.764135 | 0.963191 |
| CompactCNN end-to-end | 48 | 4924 | 8 | 0.719 | 0.719 | 0.717887 | 0.950877 |
| CompactCNN end-to-end | 49 | 4924 | 5 | 0.696 | 0.696 | 0.699727 | 0.941723 |
| CompactCNN end-to-end | 50 | 4924 | 12 | 0.796 | 0.796 | 0.797649 | 0.963372 |
| CompactCNN end-to-end | 51 | 4924 | 18 | 0.839 | 0.839 | 0.839553 | 0.974352 |

## S8. Protocol amendment and integrity

The first automated OCTMNIST-128 download produced a zero-byte file. Before formal result generation, the frozen amendment retained the OCTMNIST-64 source images verified by the official MD5 digest and mapped them to the same 16 × 16 input cells, so that only the propagation grid changed. This amendment was not selected after viewing results.

**Supplementary Table S14. Pre-result protocol-amendment record.**

| key | value |
|---|---|
| timestamp_utc | 2026-07-20T10:07:10.908069+00:00 |
| formal_results_started | False |
| reason | The official OCTMNIST-128 download produced a zero-byte file. To isolate propagation-grid resolution rather than jointly change source-image resolution, both studies use the already MD5-verified OCTMNIST-64 source and the same 16x16 input modulator. |
| protocol_change.direct_grid128_dataset_image_size | 64 |
| protocol_change.propagation_grid_size | 128 |
| protocol_change.failed_zero_byte_artifact_removed | True |
| scientific_effect | No endpoint, method, seed, training budget, physical aperture, detector width, or statistical rule changed. |
| software_source_sha256.run_octmnist_study.py | 229ca7e0c28bdf6a459006768d0407fe8b0e9c638ab17292e89f0fd67ff73b22 |
| software_source_sha256.run_gpu_extension_experiments.py | 3c3988ade2bfb974a3a6586586c84cce331c7a86beb6e94378971821ce5414ff |
| software_source_sha256.analyze_gpu_extension_experiments.py | 07bc15c845f9285719bbf2e71f6f2c6de57075b3450c31d5869eb0e4ecf5c35c |
| software_source_sha256.model.py | 730d00bb8dc2a172126f9382cce8e3fb78bc43f40cc9228dc29a48512d7f5299 |
| software_source_sha256.propagation.py | 9cf3844243017629694a735829692433655f2279bb30ec21c164096b8b53a106 |
| last_updated_utc | 2026-07-20T10:11:19.708580+00:00 |
| batch_size_clarification | 64 |
| batch_size_reason | Matches the existing physical main experiment; larger GPU-only pilot batches were rejected to avoid changing optimization dynamics. |

Complete seed-level, wavelength-shift-level, and corruption-level data and file digests are in Source_Data.xlsx. Principal sheets include GPU128_RunLevel, GPU128_PairedEffects, PhysicalDomain_RunLevel, PhysicalDomain_Paired, GPU_Environment, GPU_IntegrityAudit, and Protocol_Amendment.