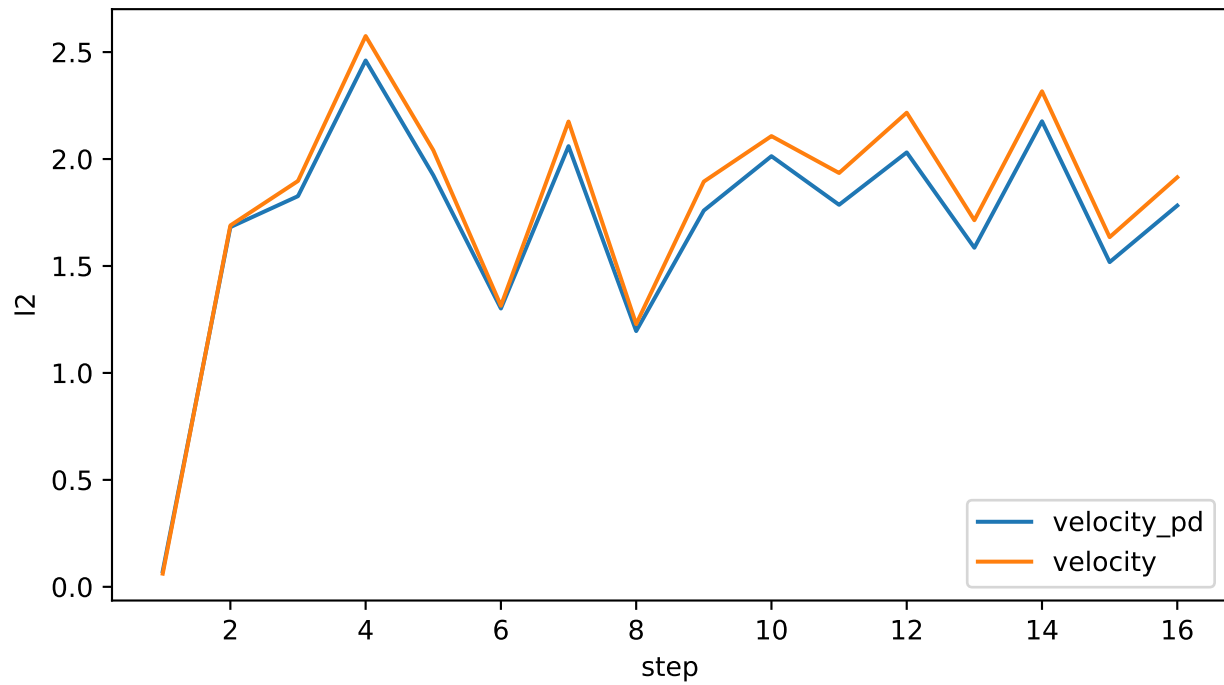


Evaluation Report

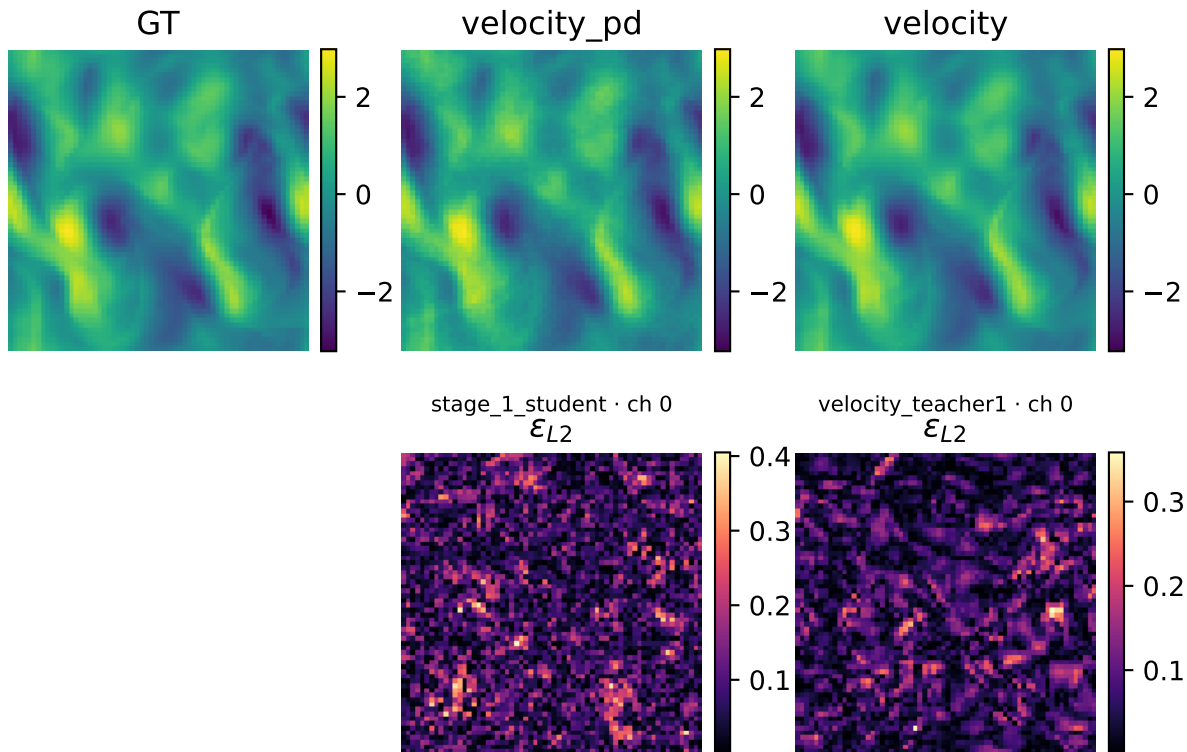
Models: velocity_2d, velocity_1d
ns2d | 2025_09_11T17_57_26

mode	model	metric	mean	median	std	count
one_step	velocity	l2	0.06188	0.05354	0.0292	32
one_step	velocity_pd	l2	0.07379	0.06867	0.03031	32
rollout	velocity	l2	1.795	1.777	0.7854	512
rollout	velocity_pd	l2	1.698	1.64	0.7483	512

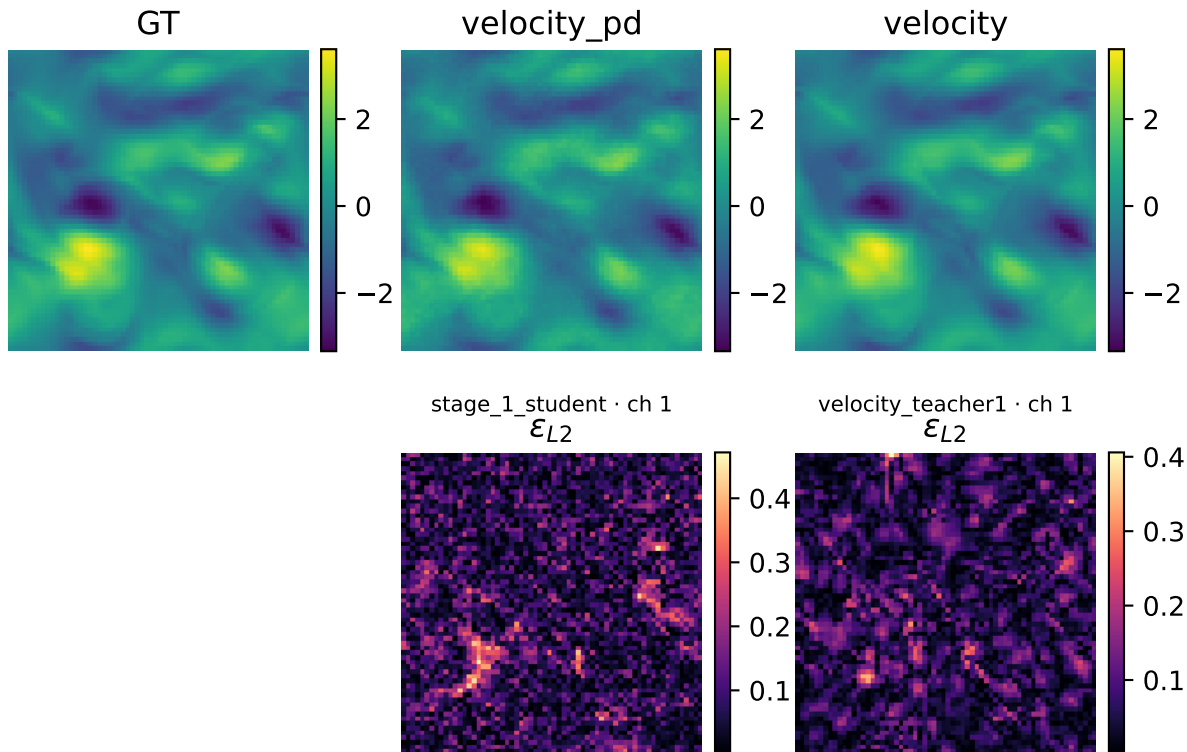
Rollout error curves (l2)



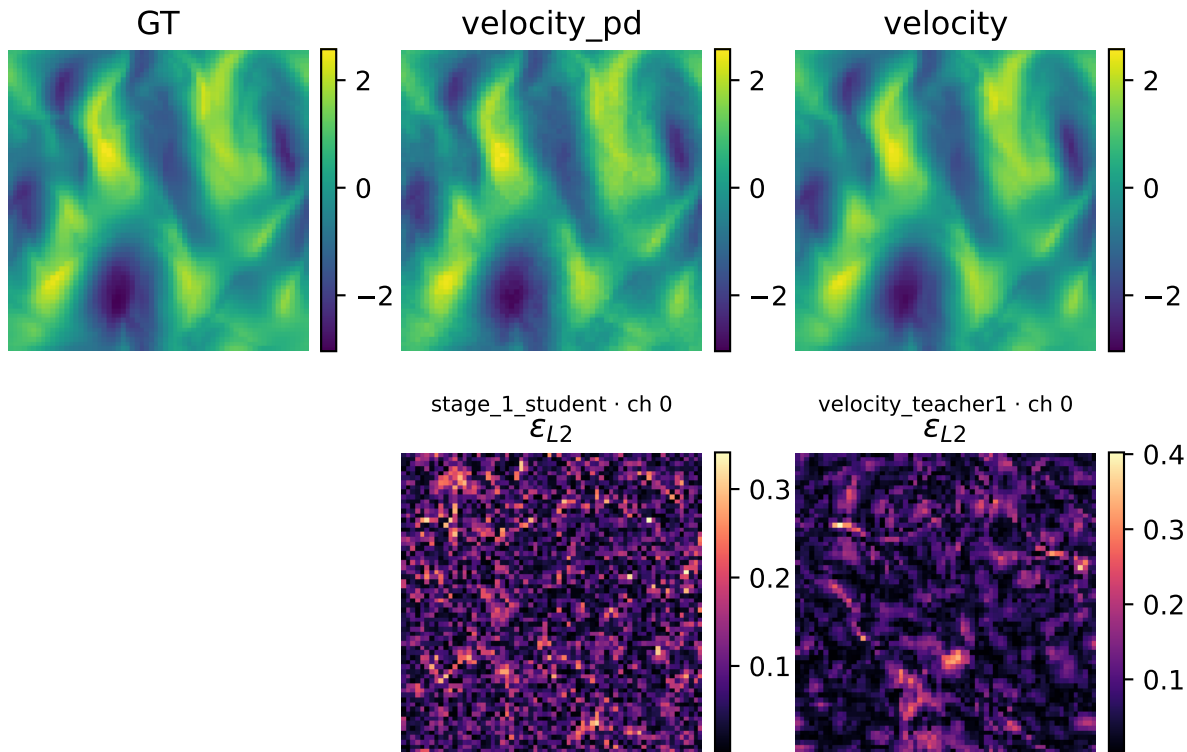
One-Step Snapshot #0 — ch 0



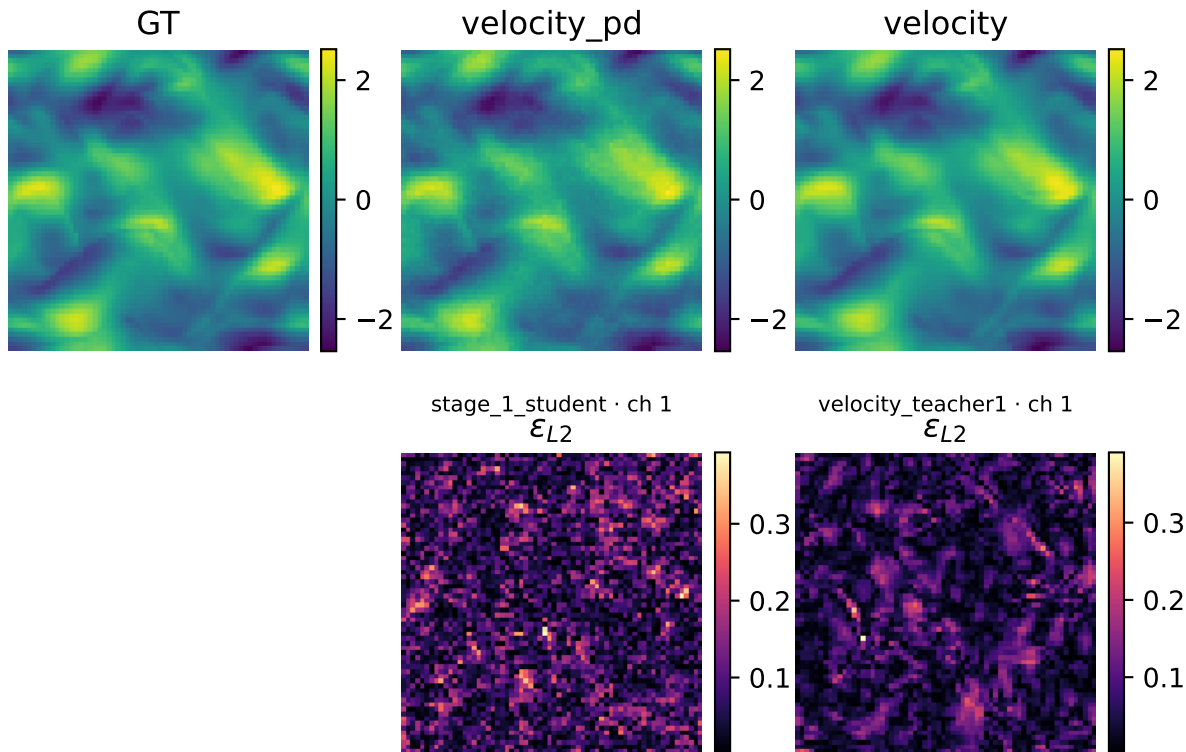
One-Step Snapshot #0 — ch 1



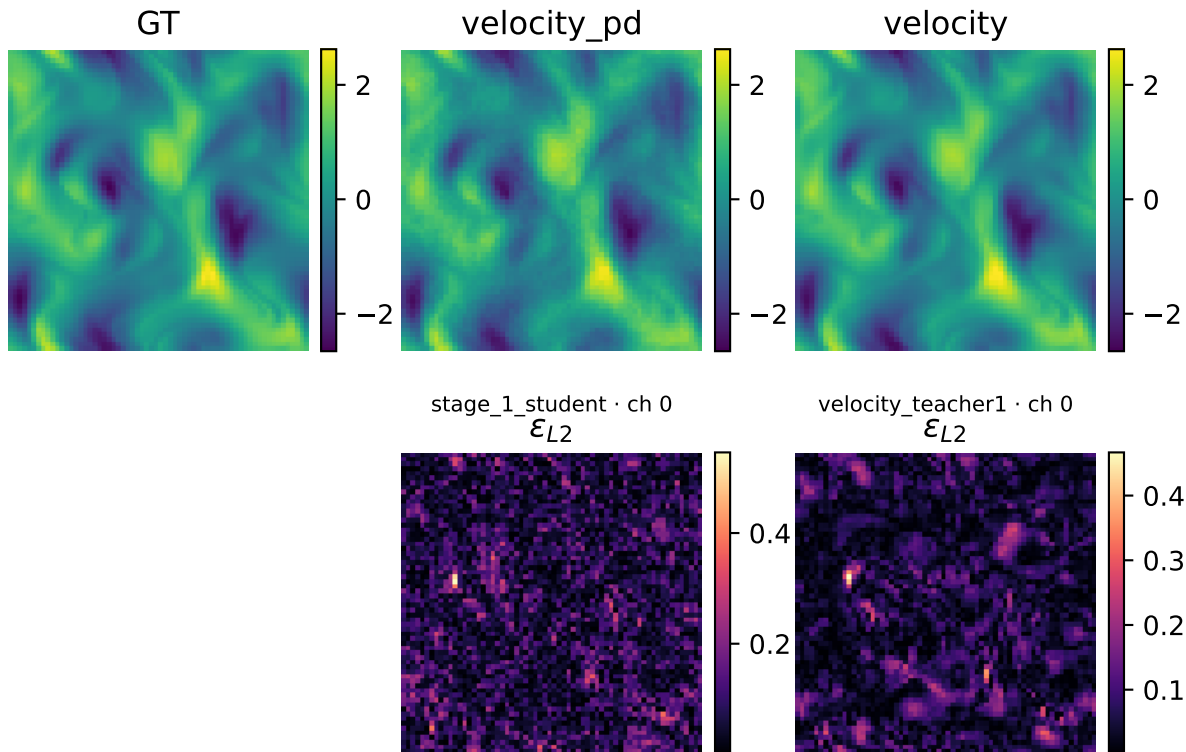
One-Step Snapshot #1 — ch 0



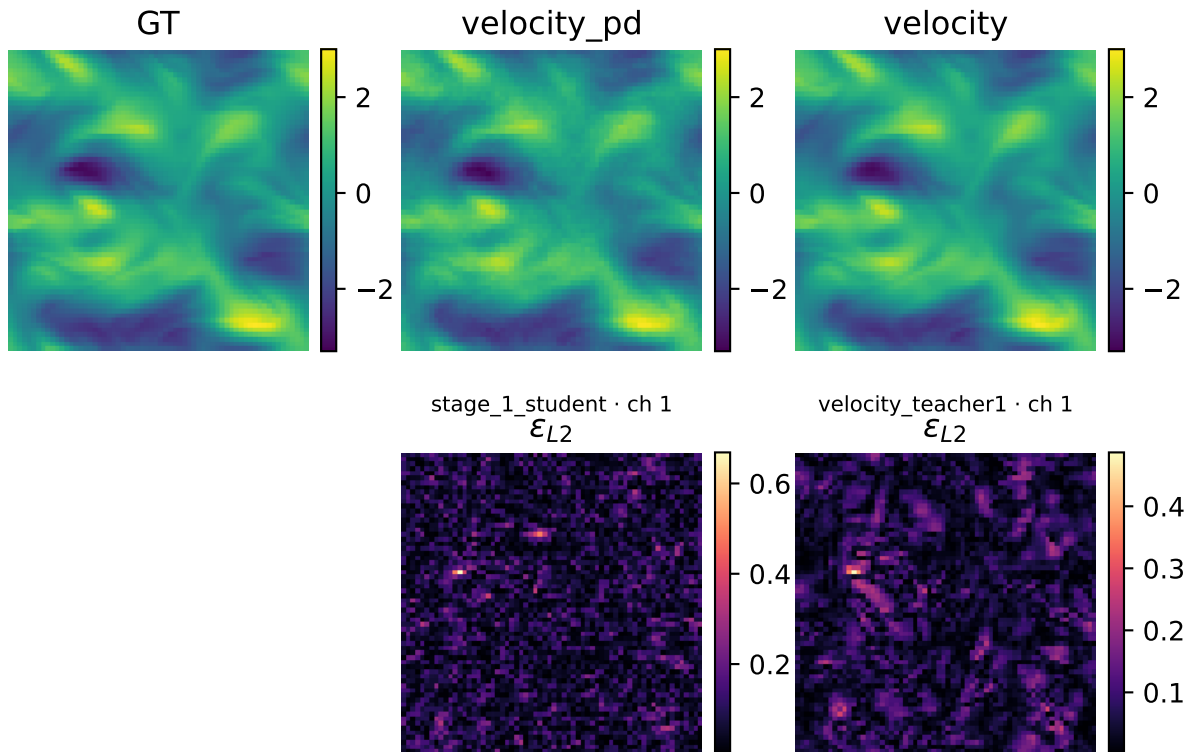
One-Step Snapshot #1 — ch 1



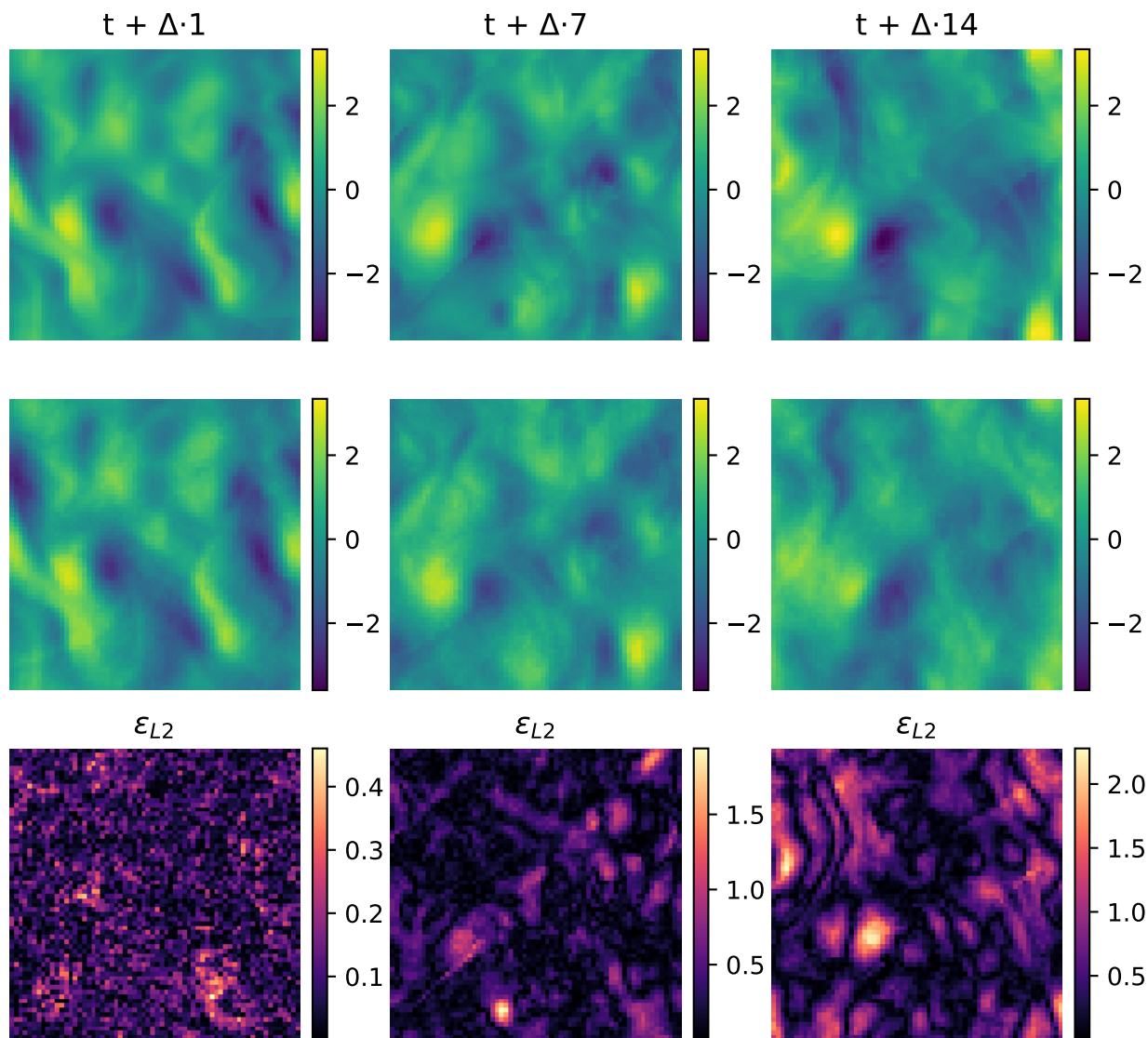
One-Step Snapshot #2 — ch 0



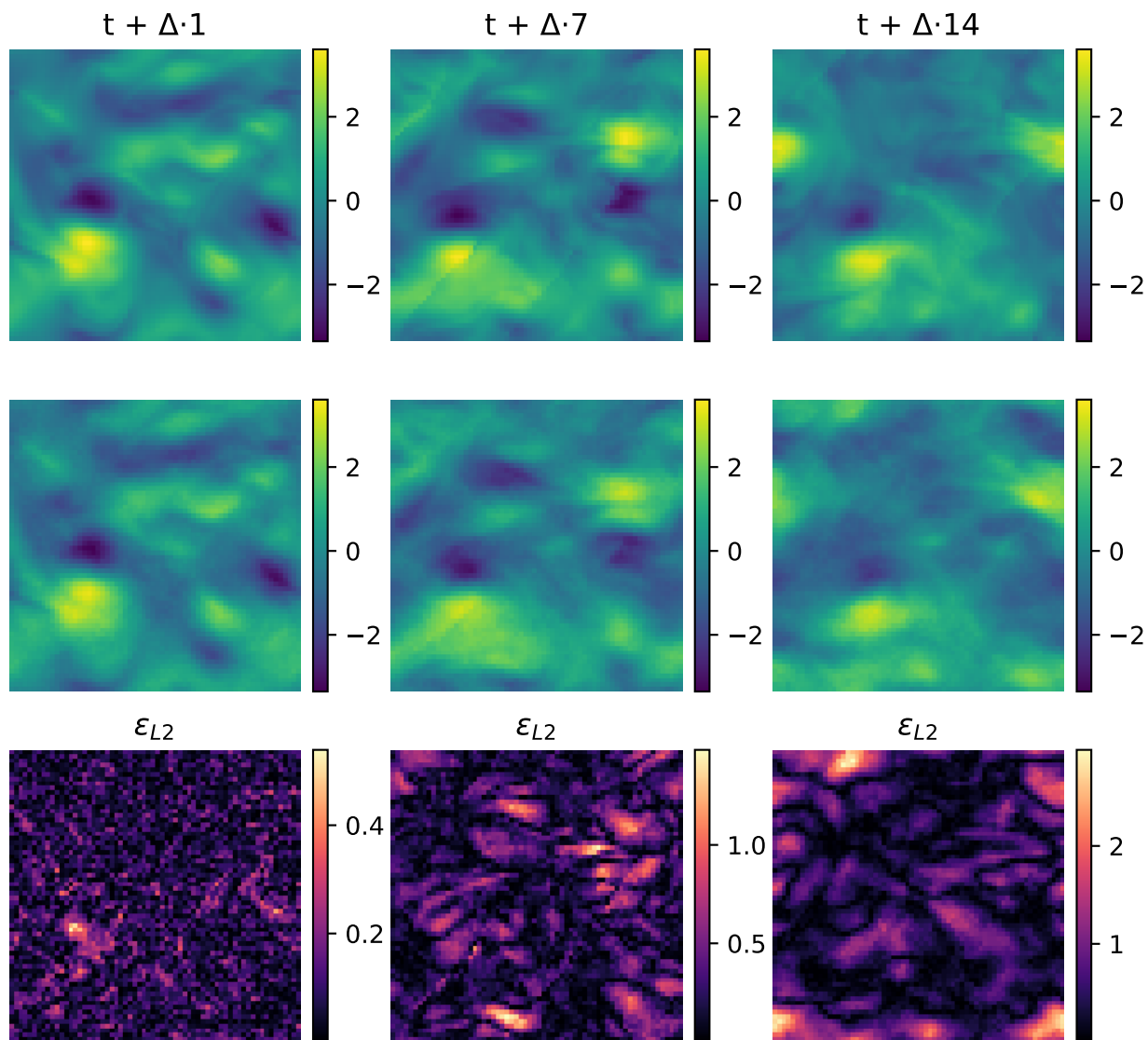
One-Step Snapshot #2 — ch 1



Rollout Snapshot #0 — velocity_pd — ch 0 · stage_1_student

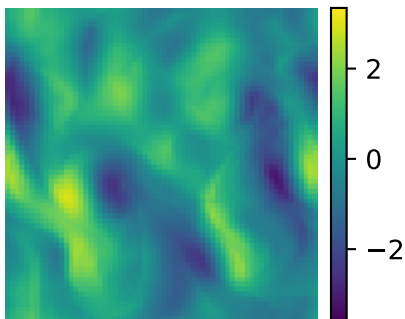


Rollout Snapshot #0 — velocity_pd — ch 1 · stage_1_student

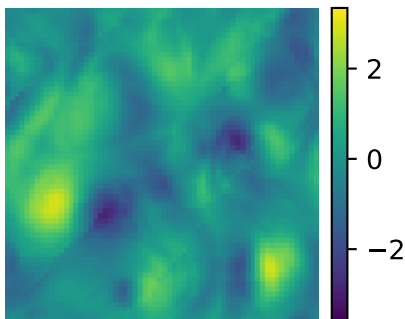


Rollout Snapshot #0 — velocity — ch 0 · velocity_teacher1

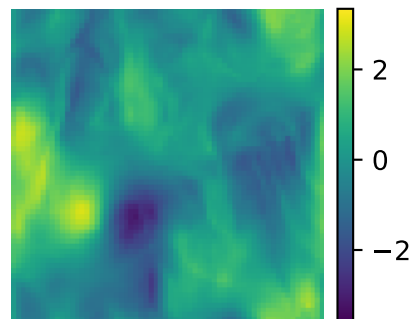
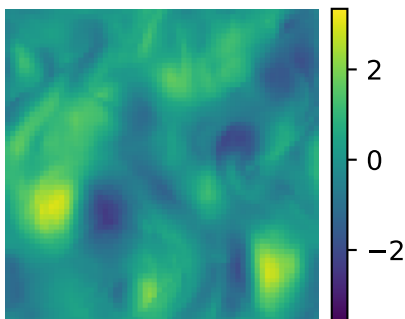
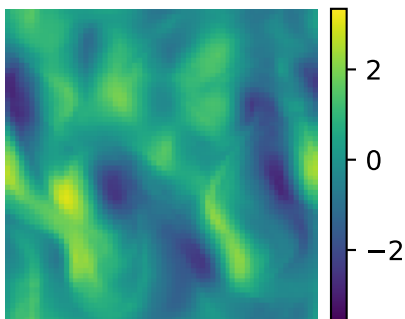
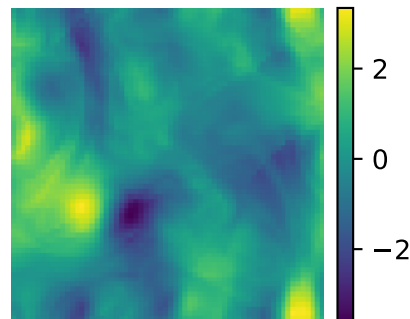
$t + \Delta \cdot 1$



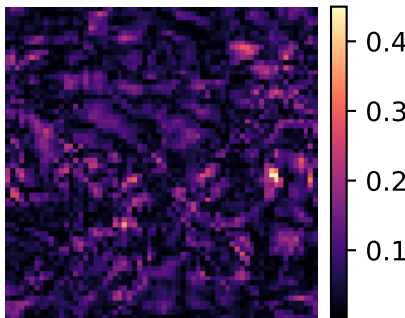
$t + \Delta \cdot 7$



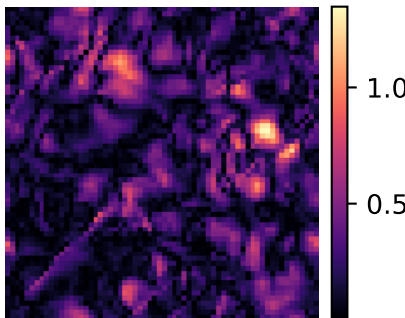
$t + \Delta \cdot 14$



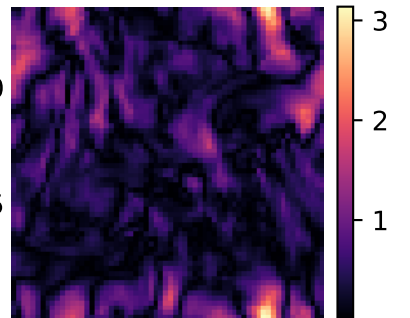
ε_{L2}



ε_{L2}

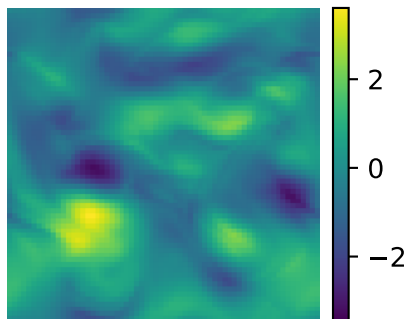


ε_{L2}

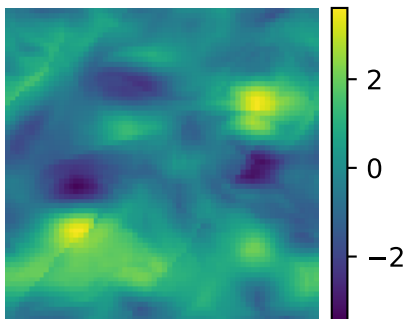


Rollout Snapshot #0 — velocity — ch 1 · velocity_teacher1

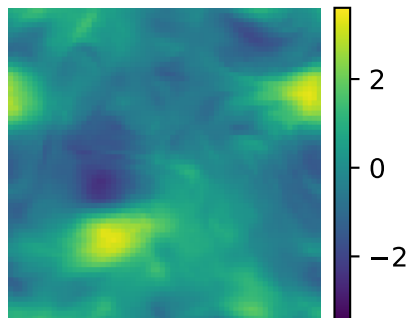
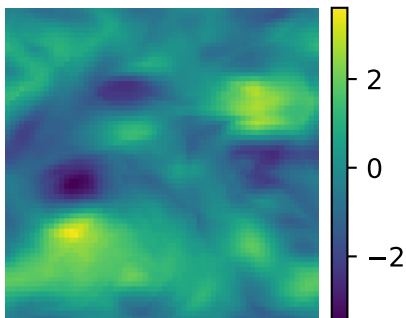
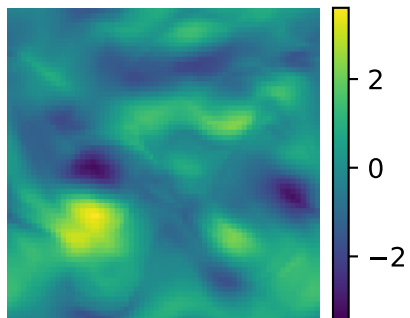
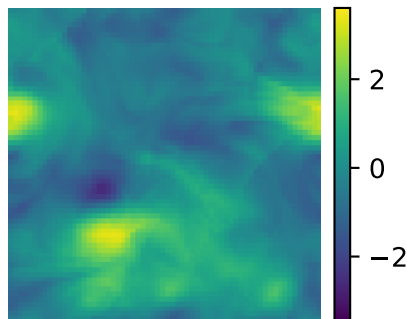
$t + \Delta \cdot 1$



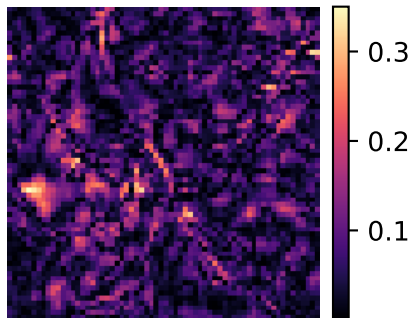
$t + \Delta \cdot 7$



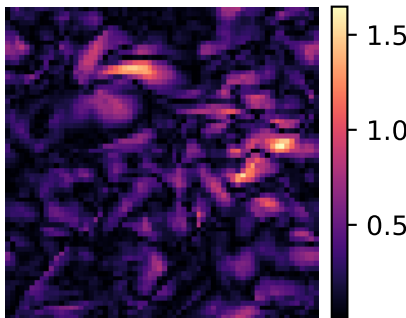
$t + \Delta \cdot 14$



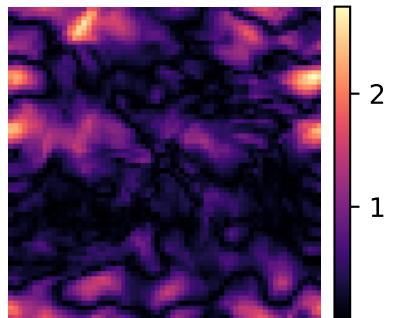
ε_{L2}



ε_{L2}

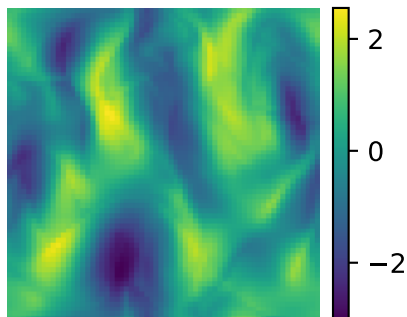


ε_{L2}

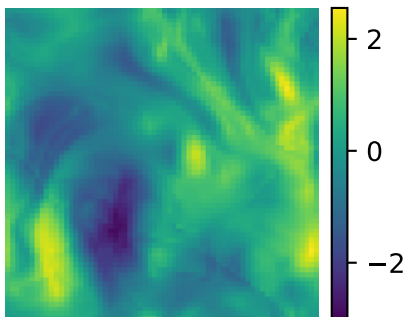


Rollout Snapshot #1 — velocity_pd — ch 0 · stage_1_student

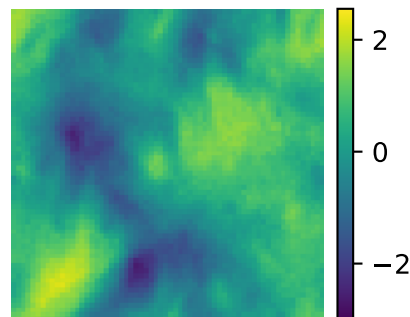
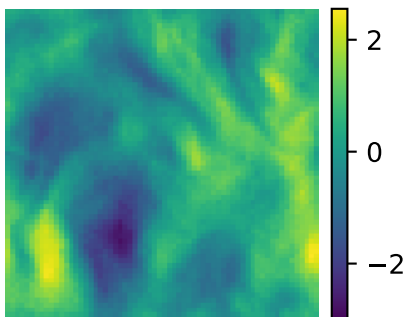
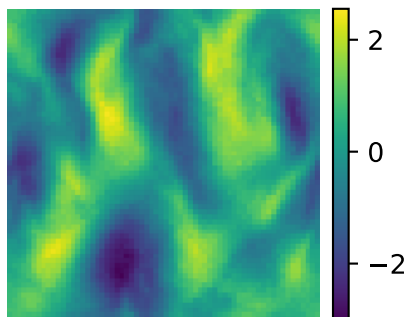
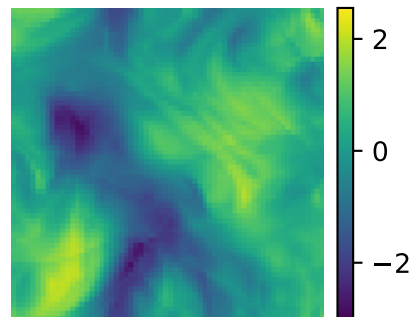
$t + \Delta \cdot 1$



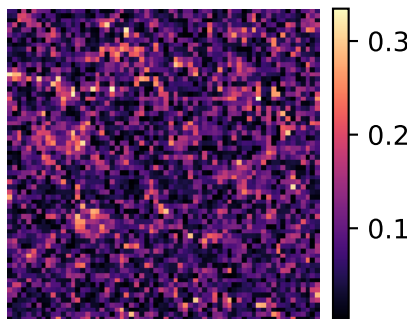
$t + \Delta \cdot 7$



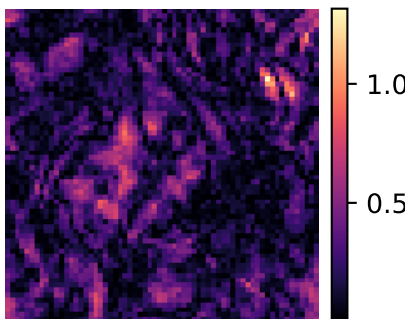
$t + \Delta \cdot 14$



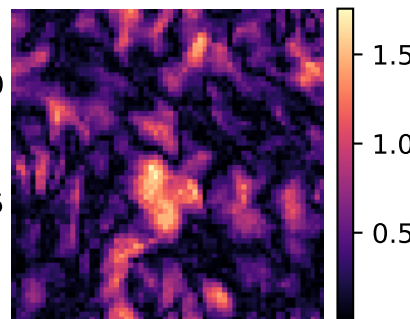
ε_{L2}



ε_{L2}

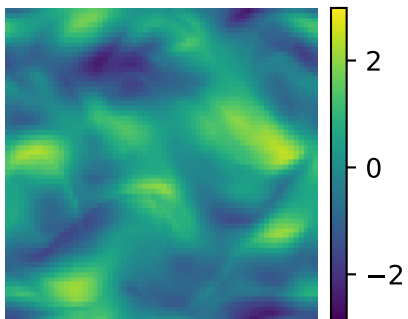


ε_{L2}

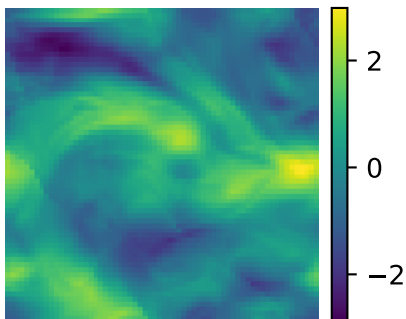


Rollout Snapshot #1 — velocity_pd — ch 1 · stage_1_student

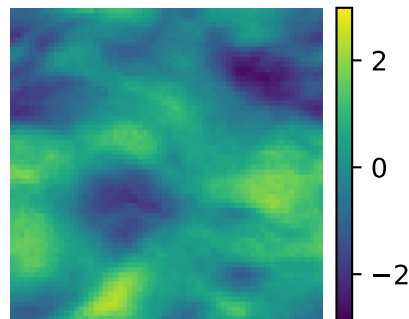
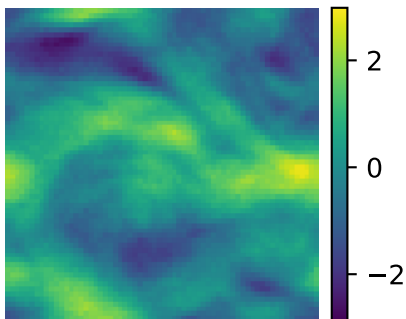
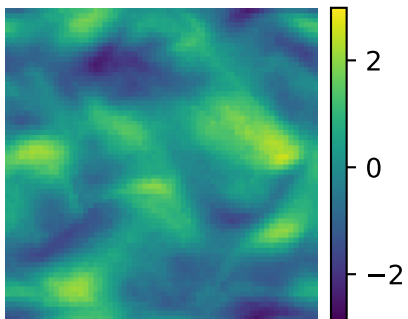
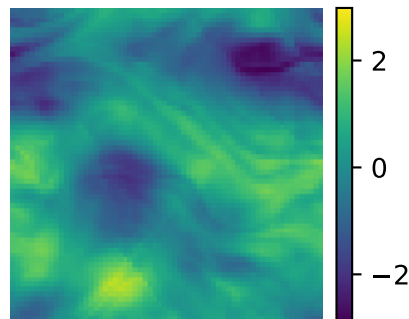
$t + \Delta \cdot 1$



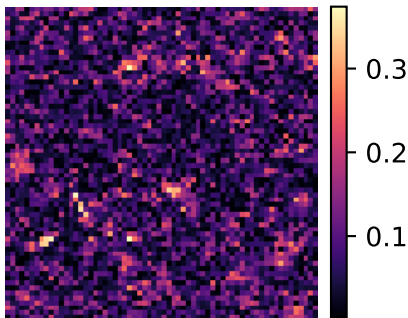
$t + \Delta \cdot 7$



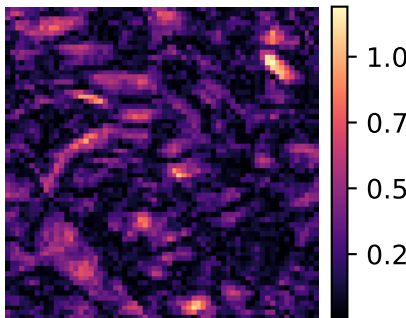
$t + \Delta \cdot 14$



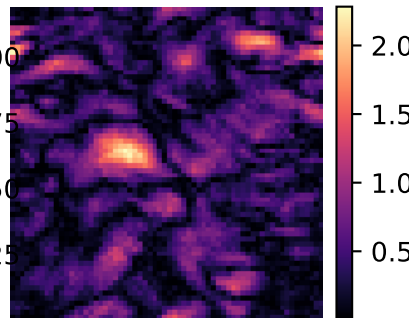
ε_{L2}



ε_{L2}

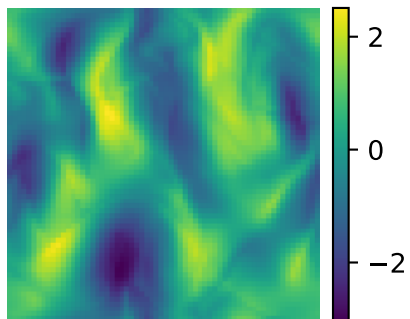


ε_{L2}

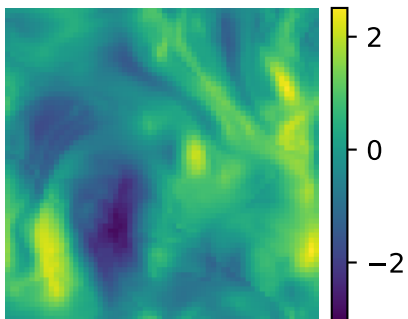


Rollout Snapshot #1 — velocity — ch 0 · velocity_teacher1

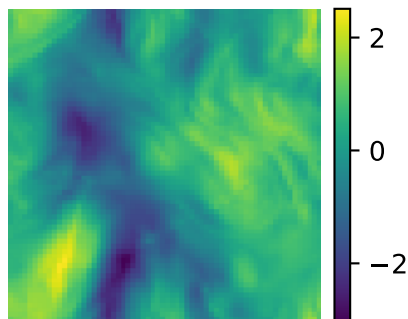
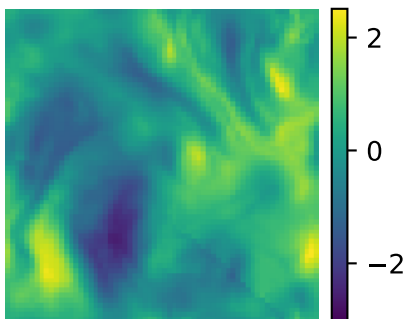
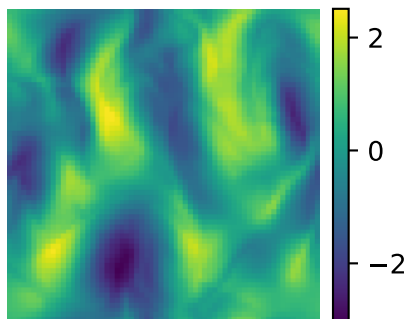
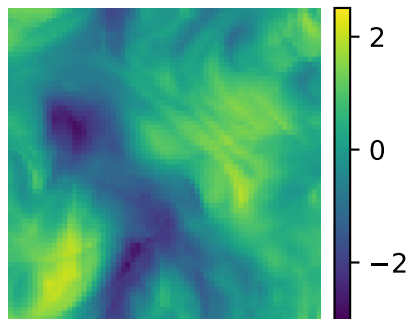
$t + \Delta \cdot 1$



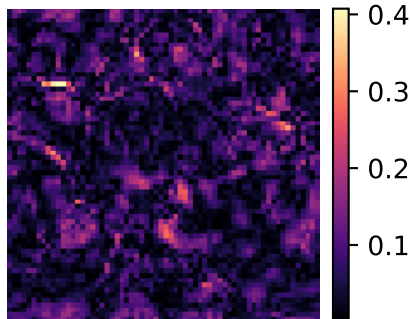
$t + \Delta \cdot 7$



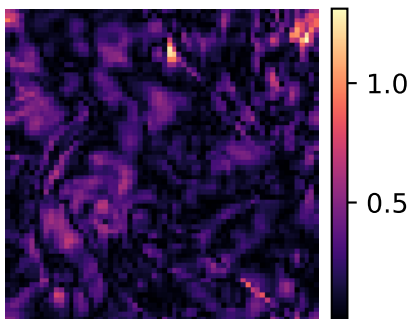
$t + \Delta \cdot 14$



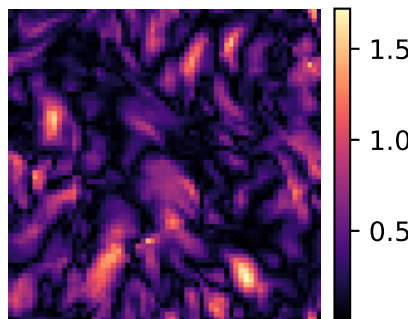
ϵ_{L2}



ϵ_{L2}

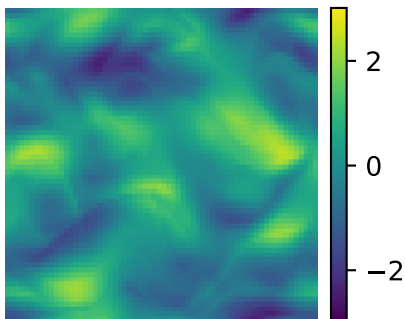


ϵ_{L2}

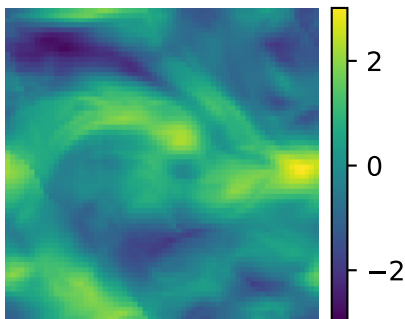


Rollout Snapshot #1 — velocity — ch 1 · velocity_teacher1

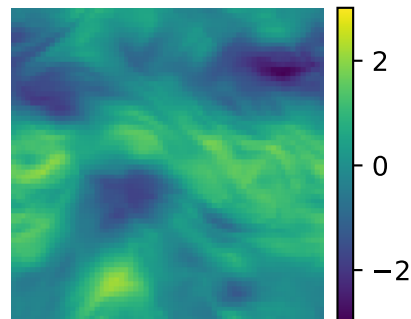
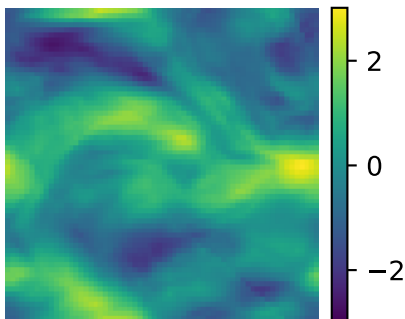
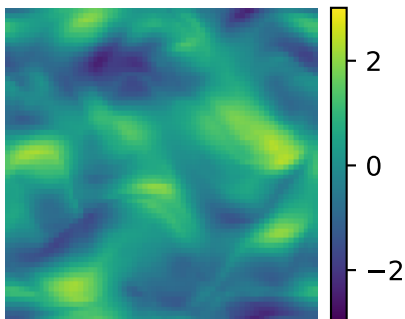
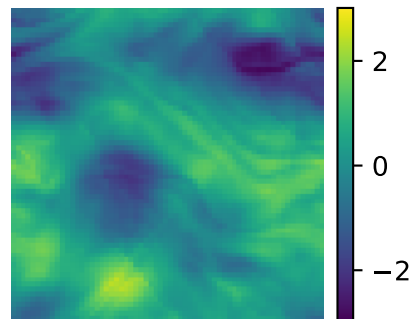
$t + \Delta \cdot 1$



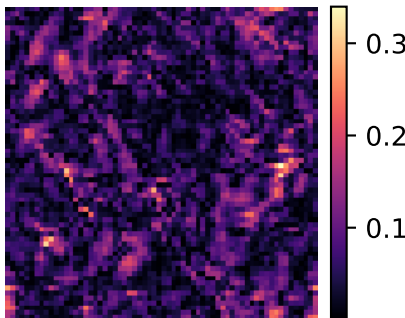
$t + \Delta \cdot 7$



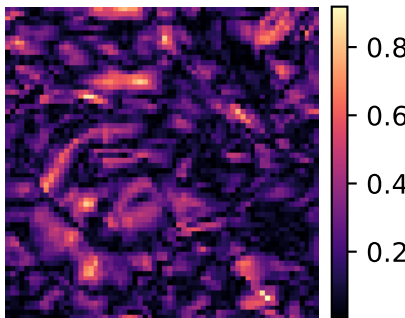
$t + \Delta \cdot 14$



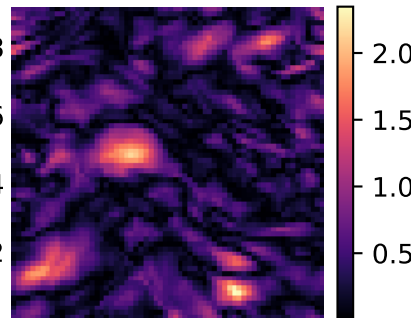
ε_{L2}



ε_{L2}

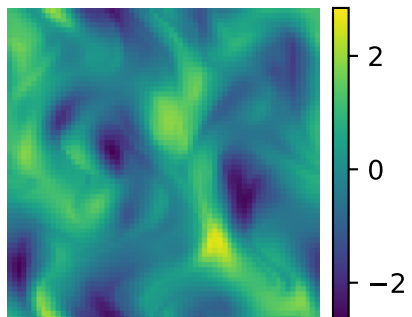


ε_{L2}

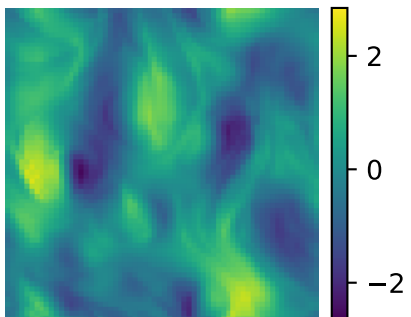


Rollout Snapshot #2 — velocity_pd — ch 0 · stage_1_student

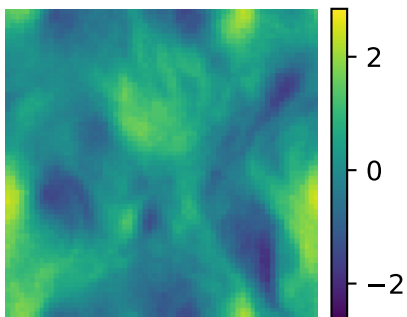
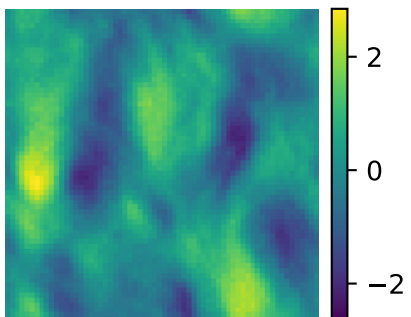
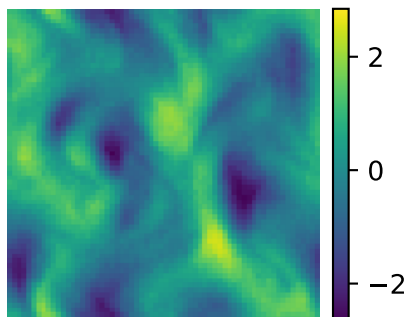
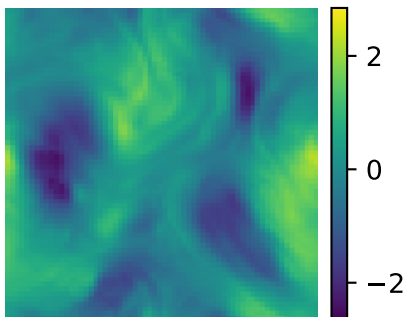
$t + \Delta \cdot 1$



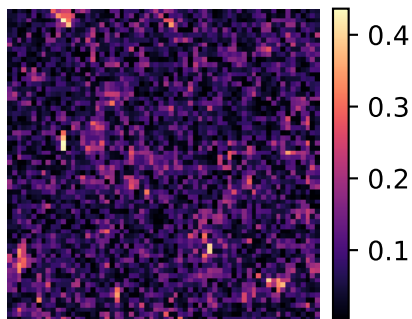
$t + \Delta \cdot 7$



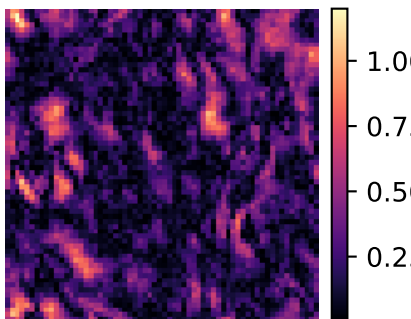
$t + \Delta \cdot 14$



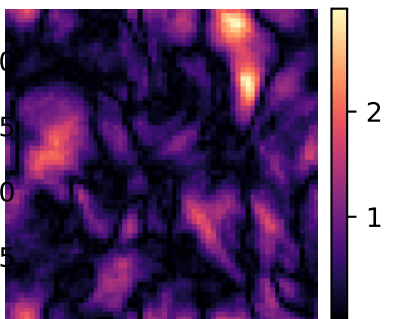
ε_{L2}



ε_{L2}

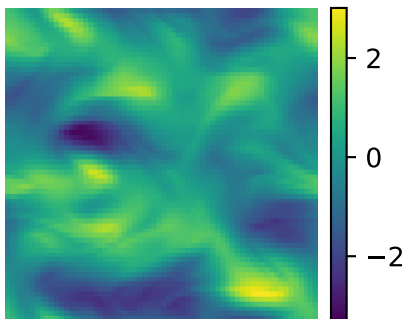


ε_{L2}

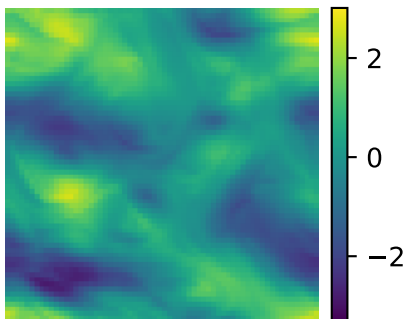


Rollout Snapshot #2 — velocity_pd — ch 1 · stage_1_student

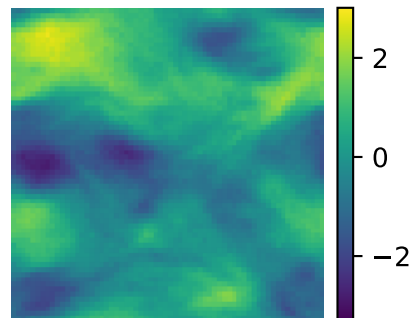
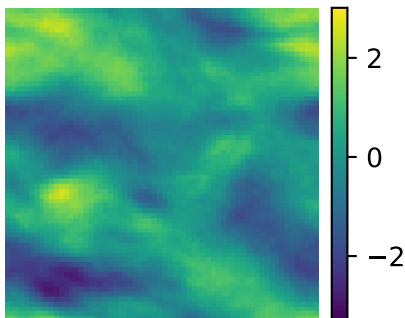
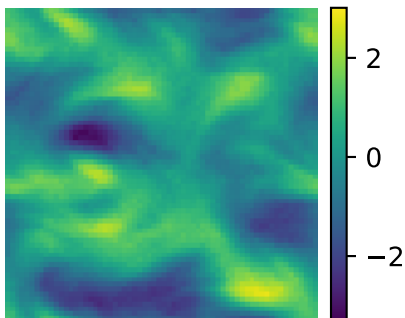
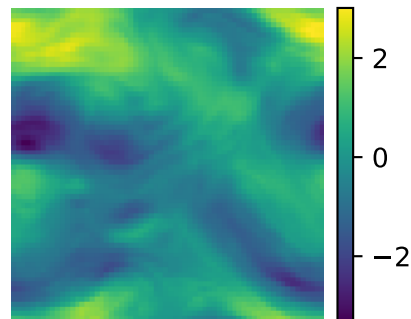
$t + \Delta \cdot 1$



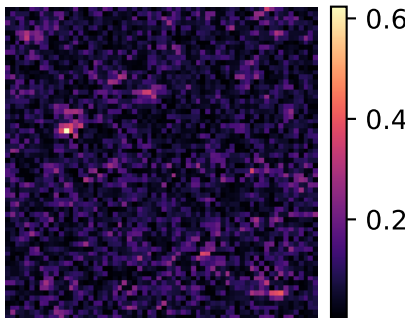
$t + \Delta \cdot 7$



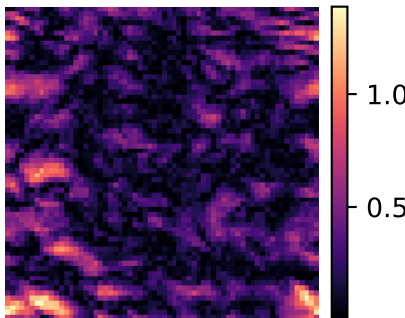
$t + \Delta \cdot 14$



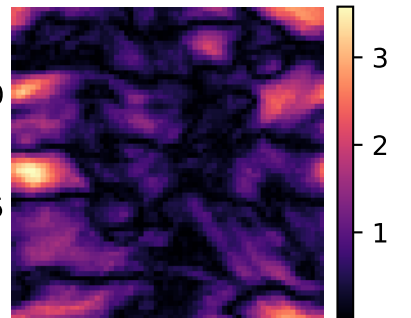
ε_{L2}



ε_{L2}

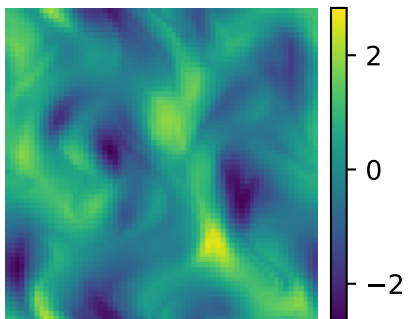


ε_{L2}

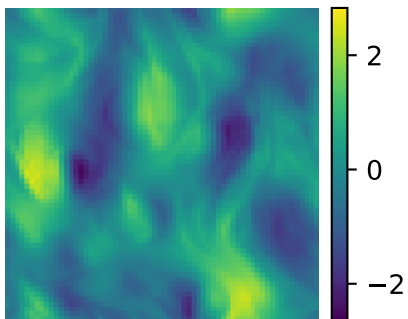


Rollout Snapshot #2 — velocity — ch 0 · velocity_teacher1

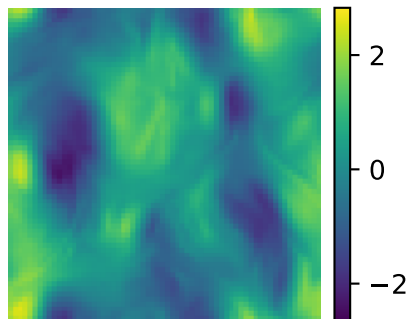
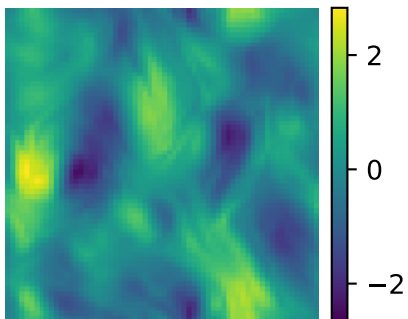
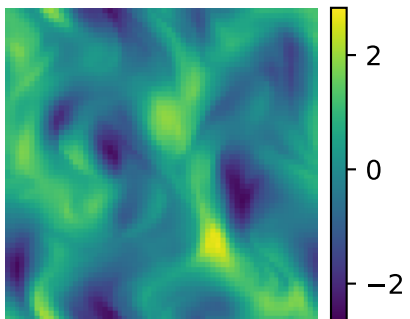
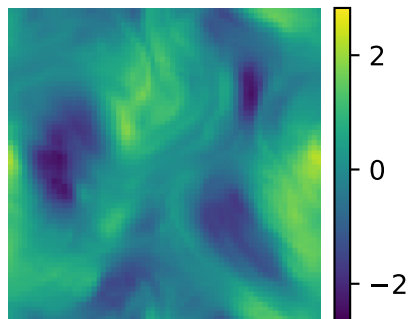
$t + \Delta \cdot 1$



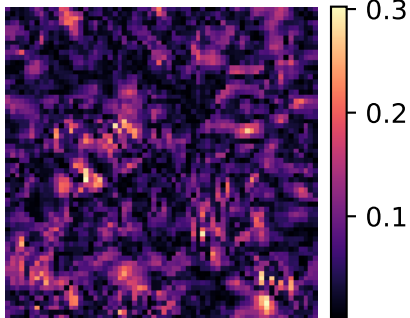
$t + \Delta \cdot 7$



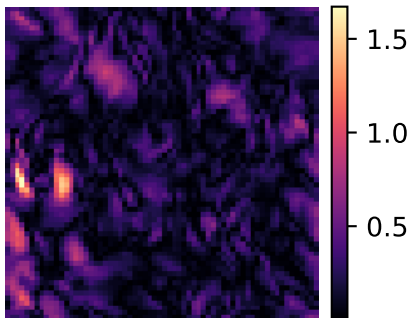
$t + \Delta \cdot 14$



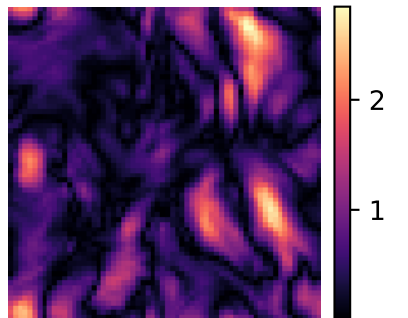
ε_{L2}



ε_{L2}

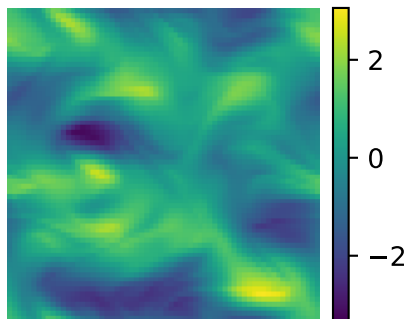


ε_{L2}

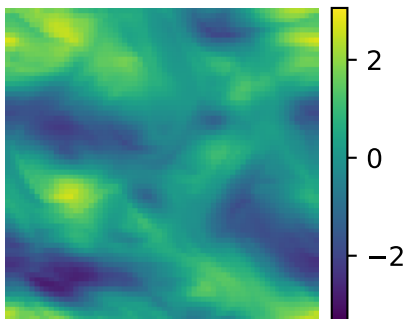


Rollout Snapshot #2 — velocity — ch 1 · velocity_teacher1

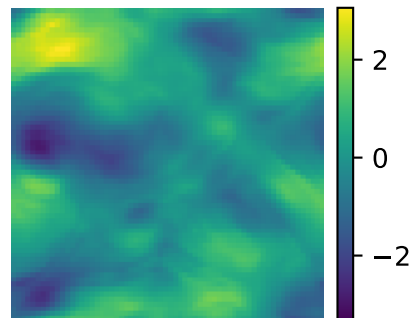
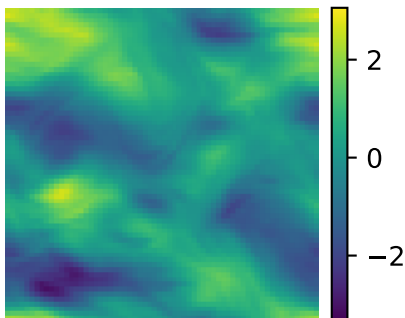
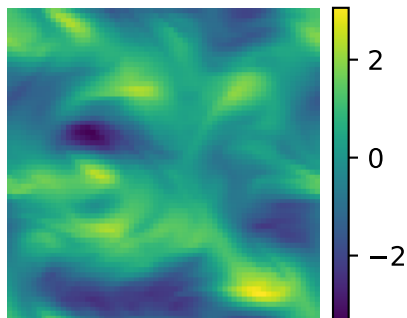
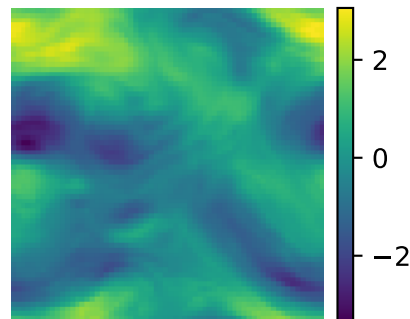
$t + \Delta \cdot 1$



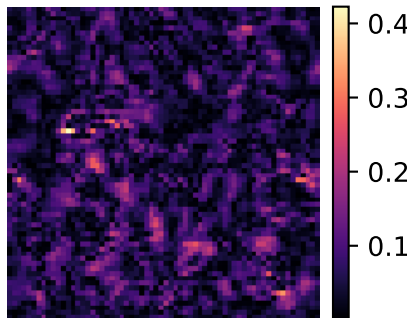
$t + \Delta \cdot 7$



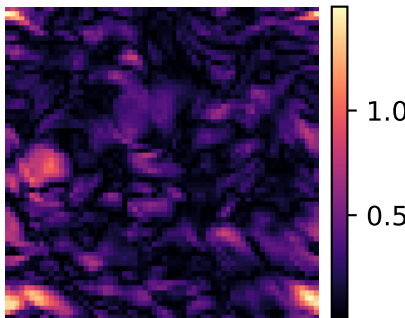
$t + \Delta \cdot 14$



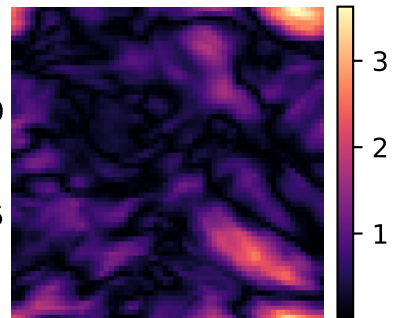
ϵ_{L2}



ϵ_{L2}



ϵ_{L2}



Evaluation Configuration

```
{
  "common": {
    "device": "cuda",
    "seed": 42,
    "samples": 32,
    "batch_size": 16
  },
  "models": [
    "velocity_pd",
    "velocity"
  ],
  "viz": {
    "enabled": true,
    "n_samples": 3,
    "channels": [
      0,
      1
    ],
    "colorbar": true
  },
  "one_step": {
    "steps": {
      "velocity": 50,
      "flow": 1,
      "velocity_pd": 25
    },
    "method": {
      "velocity": "midpoint",
      "velocity_pd": "midpoint"
    }
  },
  "rollout": {
    "length": 16,
    "latent_policy": "new_latent",
    "steps": {
      "velocity": 50,
      "flow": 1,
      "velocity_pd": 25
    },
    "method": {
      "velocity": "midpoint",
      "velocity_pd": "midpoint"
    },
    "use_sequence_gt_for_metrics": true
  },
  "metrics": {
    "enabled": [
      "l2"
    ],
    "params": {
      "sob": {
        "alpha": 1.0,
        "beta": 0.0
      }
    },
    "use_denormalized": true
  },
  "artifacts": {
    "pdf": {
      "filename": "eval_report.pdf",
      "dpi": 150,
      "max_figs_per_model": 6,
      "include_rollout_curves": true
    },
    "save_csv": true,
    "save_json": true
  },
  "experiment": "ns2d"
}
```