NTK-Guided Implicit Neural Teaching

Supplementary Material

6. 1D Audio Fitting Task

6.1. Background and Settings

We represent raw 1D audio waveforms as a continuous function $F_\theta : \mathbb{R} \rightarrow \mathbb{R}$ that maps time t to its instantaneous amplitude $a = F_\theta(t)$. For this task, we use the *test.clean* split of the LibriSpeech [42] dataset. The implicit representation is a 5×256 SIREN [62]. For NINT, we set the hyperparameters $\xi = 0.7$, $\alpha = 10$, and $\lambda = 1.0$. All strategies share the learning rate of $\eta = 1\text{e}-4$ and the batch size of $B = 20\ \%$ (except $B = 100\ \%$ for Stand.). Since EGRA [12], SoftM. [19], and Expan. [80] are born with incompatible designs for 1D audio fitting task, we compare NINT with Stand., Unif., INT [75], and EVOS [79] (with its crossover component disabled for fair comparison). Moreover, we introduce the step-wise batch size scheduler from INT [75] to produce variant strategies, which increases B as training iteration grows. For metrics, we adopt SI-SNR [21], STOI [64], PESQ [52] to evaluate audio reconstruction qualities.

6.2. Experimental Results

Quantitative Analysis. As shown in Table 6, NINT consistently achieves the best or near-best performance across all training durations and evaluation metrics. Specifically, after 3 seconds of training, where baseline strategies still produce highly degraded outputs (*e.g.*, SI-SNR: INT = 0.97 dB; EVOS = 2.18 dB), NINT clearly surpasses others by large margins. Applying the batch-size scheduler, all strategies generally improves SI-SNR and PESQ, but can slightly reduce STOI in some cases (*e.g.*, Unif.). NINT, however, benefits consistently from the scheduler without such degradation, implying its sampling distribution is more resilient to optimization dynamics shifts. These results confirm NINT's ability for rapid convergence in audio fitting tasks.

Qualitative Analysis. Fig. 6 provides qualitative validation for the 10-second regime. The spectrogram and waveform of Unif. and EVOS show blurred harmonic structures and phase inconsistencies, while NINT preserves fine-grained harmonic structures and high-frequency energy (*e.g.*, over 3 kHz), corroborating its numerical results of top PESQ and SI-SNR scores. In summary, NINT establishes a new state-of-the-art strategy for 1D audio fitting under time-constrained training, delivering both fast convergence and high perceptual quality, validated quantitatively and qualitatively.

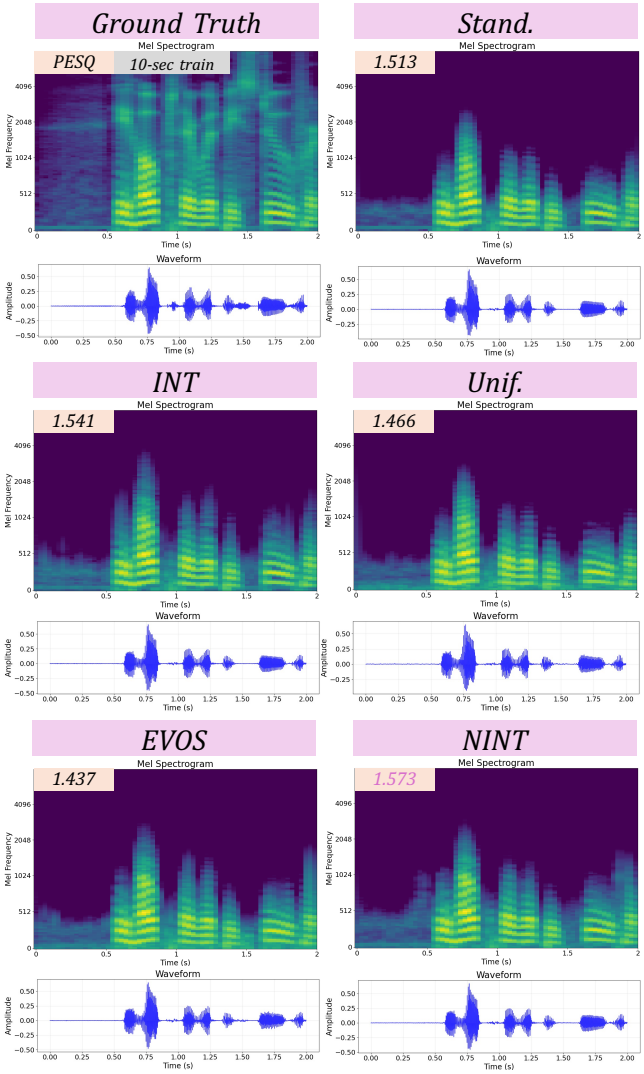

Figure 6. Visual comparison (Mel spectrogram and waveform) of audio reconstructions using different sampling strategies on the *test.clean* split of Librispeech [42] dataset, after a fixed training duration of 10 seconds.

7. 3D Shape Fitting Task

7.1. Backgound and Settings

To encode 3D shapes, we employed Signed Distance Fields (SDF), an established method in computer graphics [18]. The objective is to learn a mapping $F_\theta(\mathbf{x})$ that takes a point's coordinates (x, y, z) and outputs s, the signed distance between that point and the object's surface. Network configuration, NINT hyperparamters, and learning rate fol-

Strategy	3 seconds			10 seconds			15 seconds		
	SI-SNR↑	STOI↑	PESQ↑	SI-SNR↑	STOI↑	PESQ↑	SI-SNR↑	STOI↑	PESQ↑
Stand.	-4.786	0.507	1.098	13.054	0.698	1.513	14.455	0.703	1.570
Unif.	-4.281	0.501	1.099	11.440	0.704	1.466	12.472	0.705	1.493
INT [75]	0.968	0.590	1.175	14.053	0.694	1.541	14.948	0.714	1.675
EVOS [79]	2.178	0.563	1.107	12.014	0.697	1.437	13.438	0.695	1.547
NINT	3.842	0.614	1.200	14.280	0.702	1.573	14.803	0.715	1.834
Unif.	-3.668	0.544	1.176	11.246	0.701	1.430	13.527	0.708	1.517
INT [75]	1.919	0.605	1.214	14.201	0.695	1.603	14.584	0.715	1.751
EVOS [79]	2.371	0.618	1.138	12.680	0.695	1.490	12.999	0.716	1.681
NINT	4.988	0.637	1.217	14.286	0.701	1.609	14.977	0.719	1.836

Underline denotes step-wise batch size scheduler in INT [75].

Table 6. Performance metrics (SI-SNR↑, STOI↑, PESQ↑) at fixed training times across various sampling strategies on 1D audio fitting tasks. Purple denotes the best performance.

Strategy	500 Iters		1k Iters		2k Iters		5k Iters		10k Iters	
	IoU↑	CHD↓ (×1e-3)	IoU↑	CHD↓ (×1e-3)	IoU↑	CHD↓ (×1e-3)	IoU↑	CHD↓ (×1e-3)	IoU↑	CHD↓ (×1e-3)
Stand.	0.9545	6.353	0.9610	6.206	0.9681	6.106	0.9776	5.975	0.9811	5.942
Unif.	0.9434	6.636	0.9584	6.235	0.9629	6.139	0.9733	6.023	0.9801	5.978
INT [75]	0.9483	6.520	0.9594	6.623	0.9665	6.153	0.9749	6.111	0.9805	6.062
EVOS [79]	0.9538	6.405	0.9593	6.218	0.9640	6.123	0.9728	6.070	0.9802	6.019
NINT	0.9562	6.408	0.9630	6.221	0.9666	6.116	0.9762	6.023	0.9817	5.978
Unif.	0.9436	6.636	0.9587	6.232	0.9640	6.135	0.9740	6.016	0.9811	5.958
INT [75]	0.9483	6.518	0.9595	6.622	0.9631	6.150	0.9751	6.100	0.9805	6.020
EVOS [79]	0.9540	6.404	0.9596	6.219	0.9644	6.149	0.9733	6.059	0.9814	5.999
NINT	0.9563	6.391	0.9637	6.216	0.9670	6.138	0.9770	6.005	0.9825	5.921

Underline denotes step-wise batch size scheduler in INT [75].

Table 7. Performance metrics (IoU↑ and CHD↓) at fixed training iterations across various sampling strategies on 3D shape fitting tasks. Purple denotes the best performance.

low Sec. 6.1, while all strategies share the batch size of $B = 40\ \%$ (except $B = 100\ \%$ for Stand.). Similar to incompatibility issues as illustrated in Sec. 6.1, we compare NINT with Stand., Unif., INT [75], and EVOS [79] (crossover component disabled for compatibility), with step-wise batch size scheduler from INT [75] introduced to produce variants. Evaluation are performed on the Stanford 3D Scanning Repository [61], a well-known dataset which provides high-quality 3D scans of real-world objects and is widely used for research in computer graphics. For training set, 50,000 points are randomly sampled from the 3D shape surface on both coarse (Laplacian noise with variance 0.1) level and fine (Laplacian noise with variance 0.001) level to constitute the training set for each iteration following [27]. For metrics, we evaluate 3D reconstruction quality using: Intersection over Union (IoU), the volumetric overlap between predicted and ground-truth occupancies as a measure of global shape accuracy; and the Chamfer Distance (CHD), which is the bidirectional mean squared distance between surface point samples.

7.2. Experimental Results

Quantitative Analysis. As shown in Table 7, NINT consistently achieves the best or near-best performance across training budgets and metrics. Remarkably, NINT attains the highest IoU at every iteration threshold (500, 1k, 2k, 5k, and 10k) and the best CHD in most cases, underscoring its robust sample efficiency and balanced optimization of both global occupancy (IoU) and surface precision (CHD). Notably, when combined with the step-wise batch size scheduler, NINT shows consistent improvement over its vanilla counterpart (*e.g.*, boosting IoU from 0.9817 to

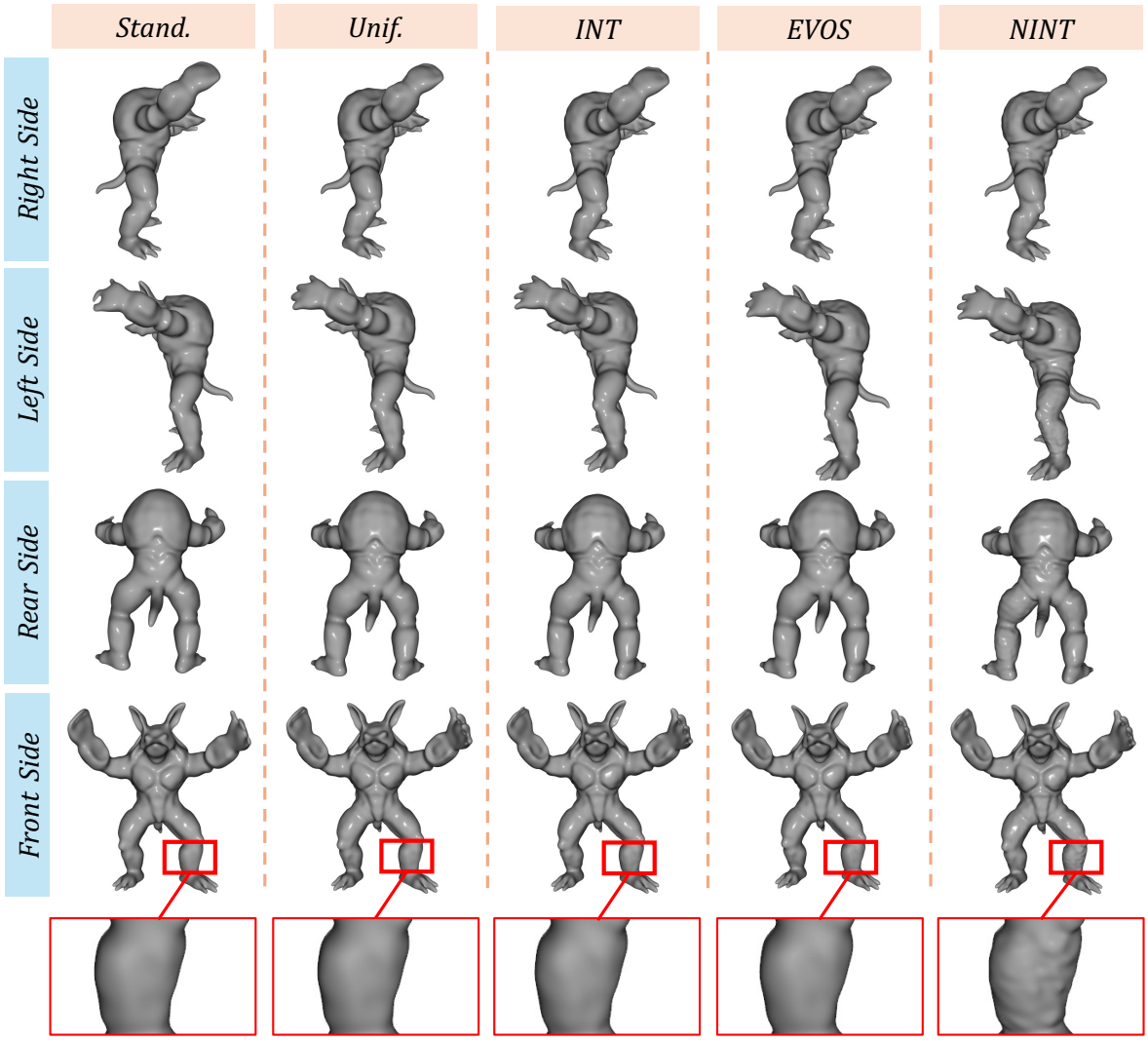

Figure 7. Visual comparison of 3D shape reconstructions using different sampling strategies on the *Armadillo* of Stanford 3D Scanning Repository [61] dataset, after a fixed training duration of 10 minutes.

0.9825) and reducing CHD from $5.978\,\mathrm{e}-3$ to $5.921\,\mathrm{e}-3$ at 10k iterations which represents the best overall results. In contrast, while schedulers benefit some baselines (*e.g.*, Unif. achieves CHD $= 5.958\,\mathrm{e}-3$), they often fail to improve both metrics simultaneously (*e.g.*, INT improves CHD slightly but lags in IoU). These results confirm that NTK-guided sampling can effectively prioritize informative regions (*e.g.*, near-surface points with high gradient variance), accelerating convergence without sacrificing geometric fidelity, making it well-suited for 3D SDF learning tasks.

Qualitative Analysis. Fig. 7 shows a visualized result of 10-minute 3D shape training on the *Armadillo* of Stanford 3D Scanning Repository [61]. Particularly, Stand. and Unif. exhibit noticeable surface noise and missing thin structures; INT and EVOS show improved topology but still suffer from bulging or over-smoothed regions. NINT, however, produces clean, high-fidelity reconstructions with sharp edges, accurate thin parts, and minimal artifacts. This corroborates the superiority of NINT's top numerical results (outstanding IoU and CHD scores), confirming that NTK's guidance effectively preserves both global shape coherence and local surface details.

8. Extended Evaluation on 2D Images

In this section, we provide extended experimental results on 2D image fitting tasks. Fig. 8 shows a visualized comparison with Stand., Unif., EVOS [79], and Expan. [80] on training multiple images from DIV2K [1] dataset for fixed 60 seconds. Across all images, NINT consistently achieves the best reconstruction quality, which is also corroborated

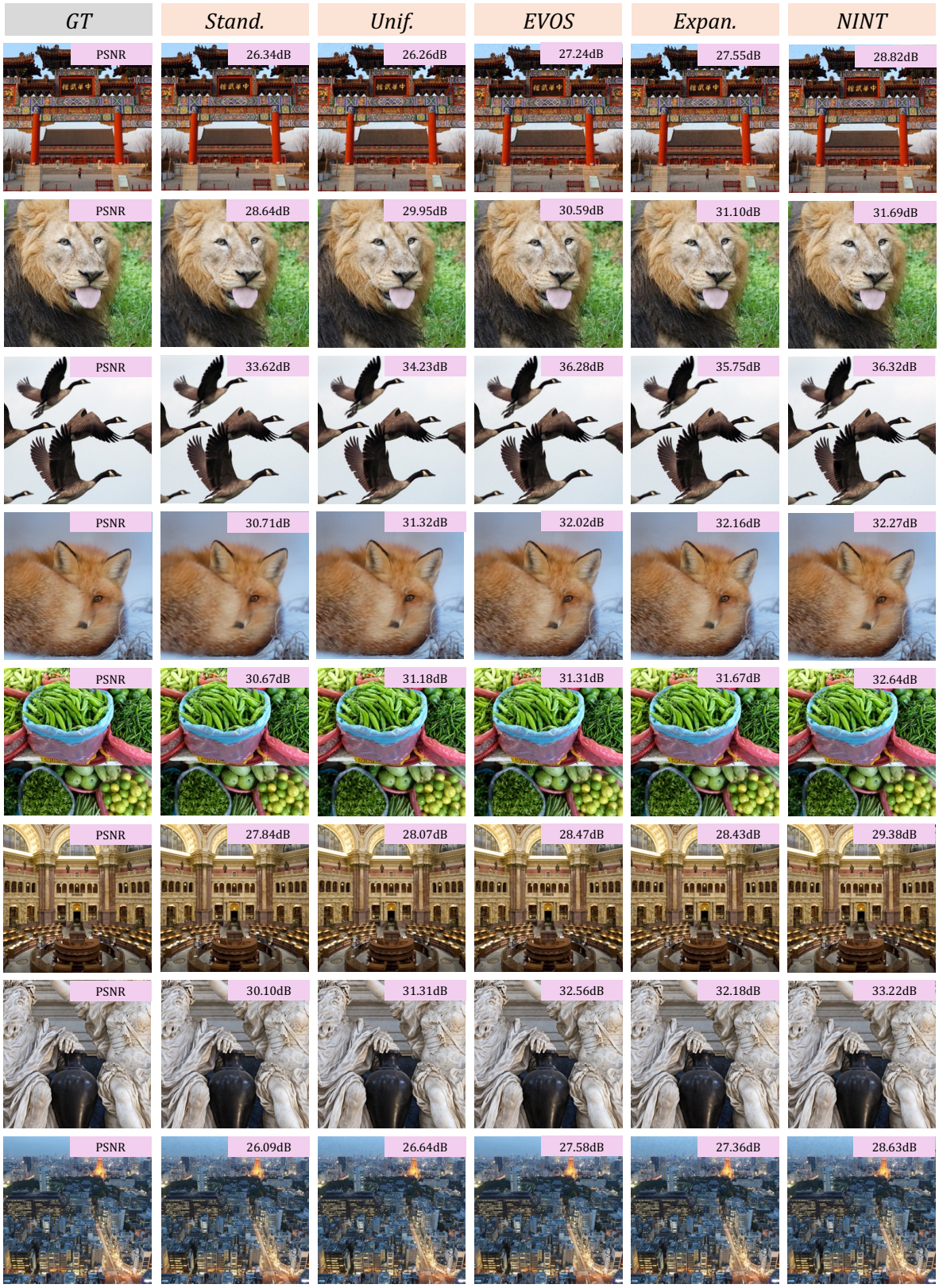

Figure 8. Visual comparison of 2D image reconstructions using different sampling strategies on DIV2K [1] dataset, after a fixed training duration of 60 seconds.

Setting	Iterations ↓						Time (s) ↓					
	PSNR		SSIM		LPIPS		PSNR		SSIM		LPIPS	
	30	35	0.8	0.9	0.2	0.1	30	35	0.8	0.9	0.2	0.1
Stand.	523	2043	420	1871	1645	2454	49.11	184.78	39.00	169.23	148.86	221.44
NINT (default)	389	1644	399	1639	1172	2082	25.09	102.88	26.00	102.56	73.51	130.21
$B = 40\,\%$	391	1636	375	1555	1319	2245	28.83	150.84	27.74	109.44	93.10	158.04
$B = 60\,\%$	451	1815	387	1633	1392	2177	35.85	140.64	30.98	126.79	108.14	168.48
$B = 80\,\%$	484	1898	421	1709	1461	2308	41.49	159.05	36.34	143.30	122.60	192.89
dense2	708	3110	608	2778	1672	3478	38.78	174.44	34.73	155.76	93.95	194.84
Incremental	380	4550	786	3360	1208	4450	24.95	153.56	26.32	183.57	76.12	236.95
Decremental	600	3082	404	3080	2160	3329	30.27	154.90	28.25	153.51	149.13	189.60
Step	380	1719	404	1644	1332	2123	25.21	115.62	26.63	110.26	88.19	182.28
Linear	394	1737	398	1641	1302	2185	26.32	118.79	26.59	111.84	87.32	151.98

Table 8. Ablation study on the impact of batch size B, sampling interval, and batch size scheduler in NINT, showing iterations and runtime (in seconds) required to reach target thresholds for PSNR, SSIM, and LPIPS on image fitting tasks. Purple : the best performance; Pale Purple : the secondary performance.

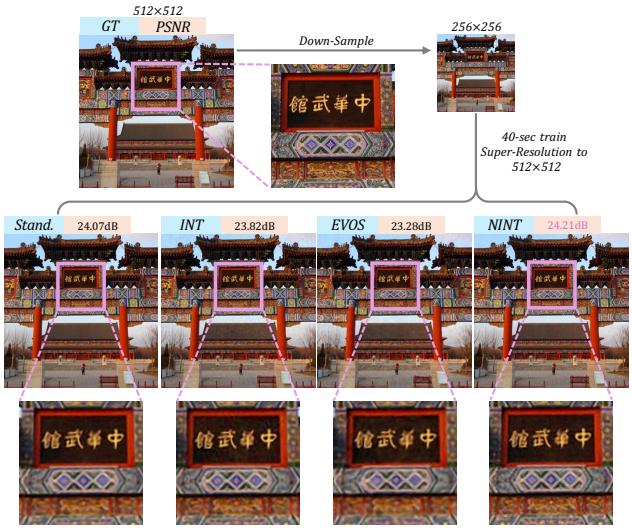

Figure 9. Visual comparison of **super-resolution** results.

by the final PSNR. This confirms the indispensability of incorporating NTK-guided sample selection into INR tasks to establish a *state-of-the-art* paradigm.

We also follow INT [75] to provide ablation studies on NINT settings including batch size B, sampling interval, and batch size scheduler. By default, NINT adopts: **(1)** batch size of $B = 20\,\%$, hence we provide comparisons of $B = 40\,\%$, $B = 60\,\%$, and $B = 80\,\%$; **(2)** sampling interval of *dense* (sample at every iteration), hence we provide comparisons of *dense2* (sample at every other iteration), *Incremental*, and *Decremental*; **(3)** constant batch size, hence we provide comparisons of *Step* batch size scheduler and *Linear* batch size scheduler.

As Table 8 shows, the default NINT configuration consistently achieves the fastest convergence in most cases, attaining the best or second-best performance in 9 out of 12 metrics. Larger batch sizes can barely improve performances in most cases, suggesting NINT's superiority of exploiting NTK's effectiveness for training acceleration. Among interval settings, *Incremental* (gradually increasing sampling frequency) yields the fastest PSNR$=$ 30 convergence (380 iterations), while the default *dense* remains most balanced overall. For batch size schedulers, the simple *Step* variant demonstrates slight improvements, yet the default constant batch size remains clearly competitive. Overall, these ablations confirm that NINT's core design, NTK-guided selection, is the primary driver of acceleration, with auxiliary choices playing secondary roles.

9. Super-resolution Task

We also demonstrate the effectiveness of NINT on **super-resolution** experiments on DIV2K dataset: downsample 512×512 → 256×256, fit INR only on low-resolution input, then evaluate high-resolution reconstruction on original ground truth. This task inherently requires the model to recover missing high-frequency details and implicitly "inpaint" structured information. As shown in Table 9 and Figure 9, NINT outperforms Stand. / INT / EVOS sampling in final PSNR/SSIM/LPIPS and reaches target quality (e.g. 20 dB PSNR) much faster.

Strategy	1000 Iterations PSNR ↑	SSIM ↑	LPIPS ↓	Time (s) ↓
Stand.	24.390	0.797	0.301	8.63
INT [75]	24.196	0.760	0.335	10.82
EVOS [79]	23.645	0.740	0.379	11.56
NINT	24.266	0.782	0.310	7.91

Table 9. **Super-resolution** performance at 1000 iterations and time to PSNR = 20 dB. Purple : the best performance.

10. Large-scale Image Fitting Task

We test NINT on 1024×1024 image fitting experiments on FFHQ dataset (Table 10 and Figure 10). NINT reaches 30 dB PSNR fastest (∼31% time reduction vs. Stand.) with visibly sharper details. The core heterogeneous structure of the NTK (varying self-leverage and coupling) remains consistent across scales, no large-image-specific patterns were observed.

Strategy	1000 Iterations PSNR ↑	SSIM ↑	LPIPS ↓	Time (s) ↓
Stand.	31.702	0.808	0.413	141.53
INT [75]	31.767	0.798	0.417	123.19
EVOS [79]	31.903	0.810	0.404	124.33
NINT	31.957	0.815	0.396	97.32

Table 10. **Large-scale** image fitting at 1000 iterations and time to PSNR = 30 dB. Purple : the best performance.

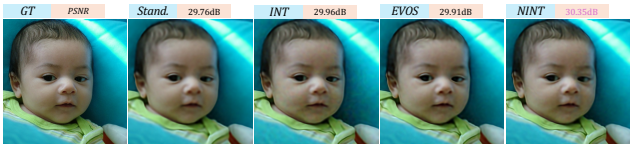

Figure 10. Visual comparison of **large-scale** image fitting results.

11. Efficient NTK Computation

A potential bottleneck in NINT is the explicit construction of the Neural Tangent Kernel (NTK) matrix $K_{\theta^t} \in \mathbb{R}^{N\times N}$. As defined in Eq. 8, calculating the full matrix is computationally heavy for high-resolution signals.

To maintain a lightweight sampling overhead, we avoid forming K_{θ^t} explicitly. Since the NTK-guided selection of $\mathcal{B}^\star$ only involves the product of the NTK and the loss gradient vector $\mathbf{g}^t$, we implement this efficiently using two automatic differentiation primitives: a Vector-Jacobian Product (VJP) followed by a Jacobian-Vector Product (JVP). Specifically, we first compute $\mathbf{v} = \mathbf{J}^\top \mathbf{g}^t$, where $\mathbf{J}$ is the Jacobian of the model outputs w.r.t. parameters, and subsequently compute the scores $\mathbf{w} = \mathbf{J}\mathbf{v} = \mathbf{J}(\mathbf{J}^\top \mathbf{g}^t) = K_{\theta^t}\mathbf{g}^t$.

In practice, it is found that an NTK scoring typically takes 3.2 ms wall-clock time (3.6% of total selection) on an NVIDIA RTX 4090, with Kodak dataset (512×512 pixels) and 5×256 SIREN. This underscores NINT's capacity to leverage negligible extra computation into significant INR training acceleration.