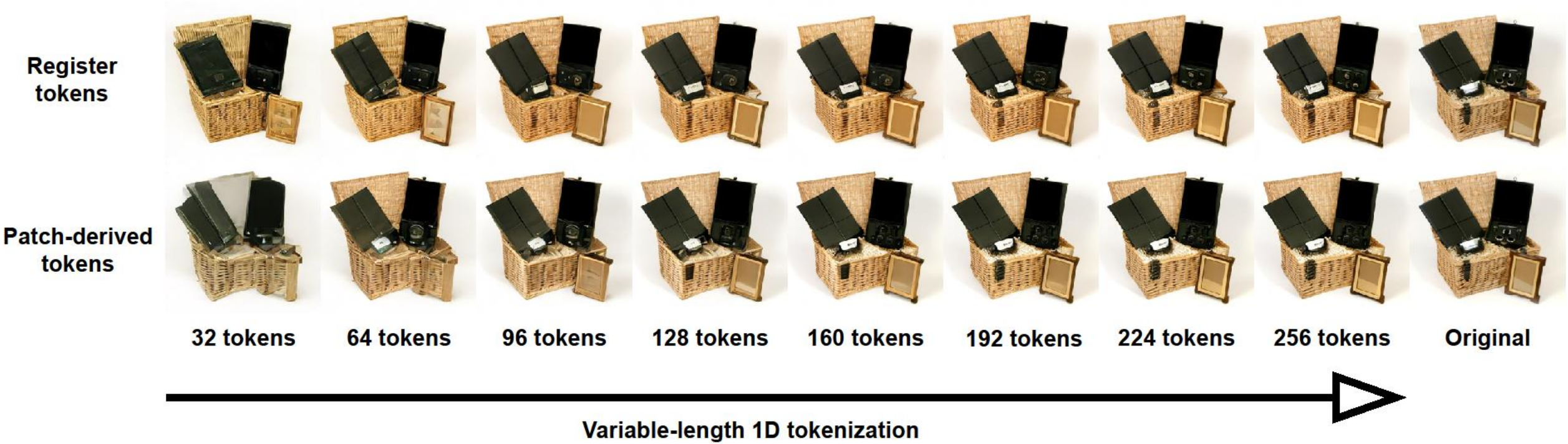
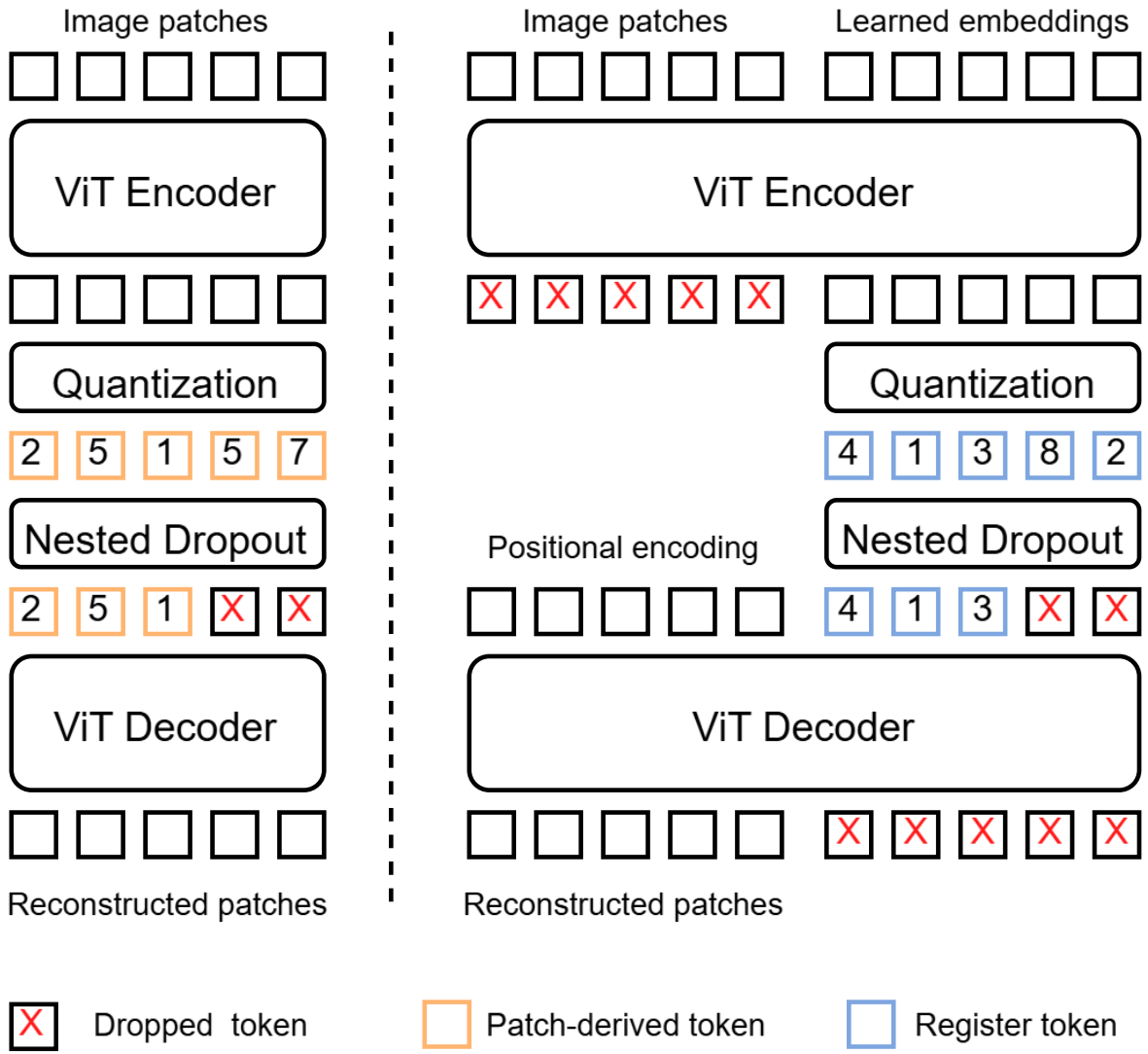
Patch-Derived or Register Tokens? Comparing Variable-Length Image Tokenizers

Philipp Schenkel, Tobias Schürmann and Ömer Şahin Taş

Variable-Length Image Tokenizers



Architecture Overview



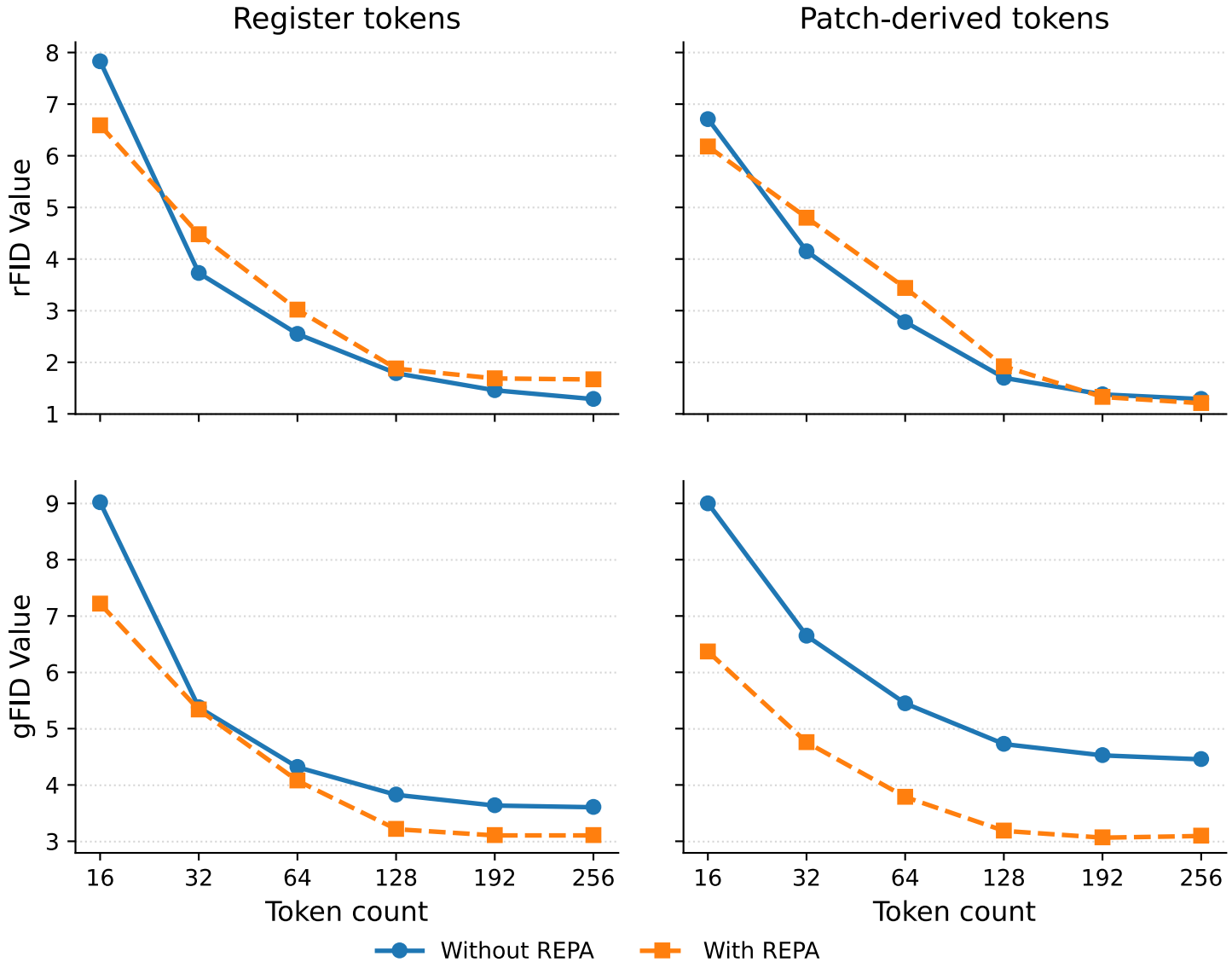
Throughput Comparison

Token type	Encoder throughput	Decoder throughput				
		1	32	64	128	256
Patch-derived	1622.9	1756.8 (constant)				
Register	901.3	1750.9	1573.9	1452.6	1217.8	932.7

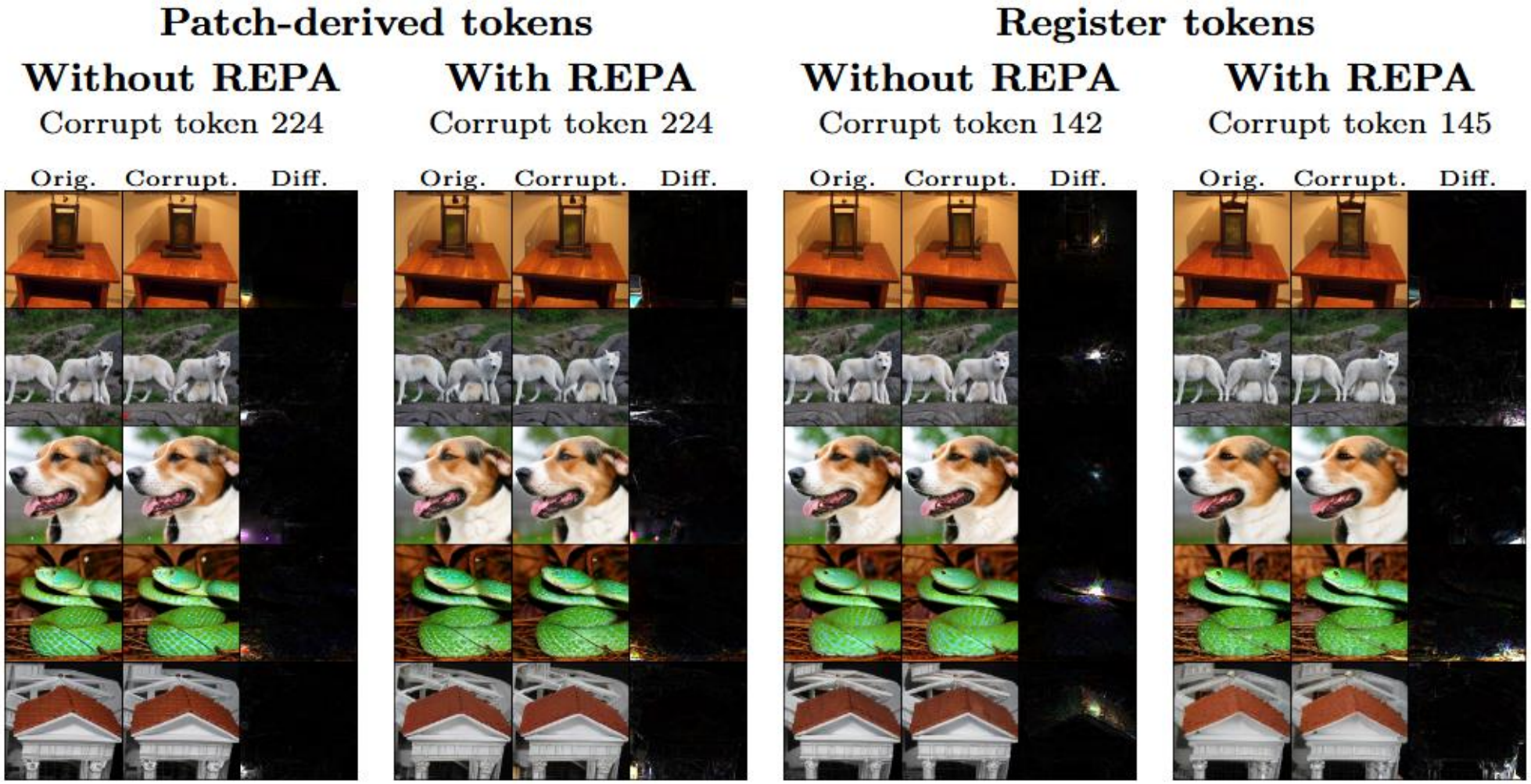
Training Details

- In total, we trained 4 tokenizers
- 2 for each architecture, once with REPA, once without REPA
- For each tokenizers, we trained several autoregressive generative models

Evaluation


Register tokens
Patch-derived tokens
rFID Value
gFID Value
Token count
Token count
16
32
64
128
192
256
Without REPA
With REPA

Corruption Analysis



Summary

- Patch-derived tokens have higher throughput compared to register tokens.
- Without REPA, register tokens yield superior performance on downstream generative tasks.
- With REPA, patch-derived tokens close this gap.
- At full token budget, many tokens exhibit local spatial influence.