

Difúzní modely (Stable Diffusion)

Difúzní modely jsou třídou generativních modelů, které se staly standardem pro generování obrazu po roce 2021. Na rozdíl od **GAN** nebo **VAE** nepracují s přímým soubojem sítí, ale využívají proces postupného přidávání a následného odstraňování šumu.

Stojí za populárními nástroji jako **Stable Diffusion**, **Midjourney**, **DALL-E 3** nebo **Imagen**.

1. Princip fungování: Destrukce a rekonstrukce

Proces fungování difúzních modelů se dělí na dvě fáze:

A. Dopředná difúze (Forward Diffusion)

Model vezme čistý obrázek a postupně k němu přidává náhodný šum (Gaussovský šum), dokud se obrázek nestane naprosto nečitelným „zrněním“. Tento proces je fixní a slouží k přípravě dat pro trénink.

B. Zpětná difúze (Reverse Diffusion)

Toto je fáze, kde probíhá učení. Neuronová síť (obvykle architektura **U-Net**) se učí odhadnout, kolik šumu bylo v daném kroku přidáno, a snaží se jej odstranit.

- Cílem je „vyčistit“ šum tak, aby se postupně objevil smysluplný obraz.
- Při generování pak modelu stačí dát čistý náhodný šum a on z něj „vysochá“ nový obrázek.

[Image showing the process from clear image to noise and back to generated image]

2. Stable Diffusion a Latentní difúze

Hlavním problémem původních difúzních modelů byla extrémní náročnost na výpočetní výkon (pracovaly s každým pixelům). **Stable Diffusion** (vyvinutý Stability AI a LMU Munich) tento problém vyřešil pomocí **Latentní difúze**.

- **Latentní prostor:** Místo aby model pracoval s pixely (např. 512×512), pracuje s komprimovanou reprezentací v tzv. latentním prostoru (podobně jako **autoenkodéry**).
- **Efektivita:** Díky tomu lze Stable Diffusion spustit i na běžných domácích grafických kartách (GPU).

3. Text-to-Image (Spojení s textem)

Aby model věděl, co má generovat (např. „kočka v klobouku“), využívá mechanismus **Conditioning**

(podmiňování).

1. **Textový enkodér (CLIP):** Převeďte textový popis (prompt) na číselný vektor.
2. **Cross-Attention:** Tento vektor je předán difúznímu modelu během procesu odšumování. Model pak při každém kroku čištění "kontroluje", zda to, co vytváří, odpovídá zadanému textu.

4. Hlavní výhody a využití

Výhoda	Popis
Kvalita a detaily	Dokáže generovat mnohem realističtější detaily než GAN.
Kontrola (ControlNet)	Možnost přesně řídit kompozici, pózu postav nebo hloubku ostrosti.
Inpainting / Outpainting	Schopnost dokreslit chybějící části obrázku nebo jej rozšířit do stran.
Open Source	Na rozdíl od DALL-E je Stable Diffusion otevřený kód, což vedlo k obrovskému rozvoji komunity.

5. Etika a právo

Difúzní modely vyvolaly velké debaty ohledně:

- Trénovací data:** Modely byly trénovány na miliardách obrázků z internetu, často bez souhlasu autorů.
- Deepfakes:** Snadná tvorba fotorealistických podvrhů známých osobností.
- Autorská práva:** Kdo vlastní obrázek vygenerovaný AI?

Související články:

- [Srovnání starších generativních modelů](#)
- [Architektura Transformer a CLIP](#)
- [Role GPU ve výpočtech AI](#)

Tagy: ai generative_models stable_diffusion deep_learning nlp

From:
<https://www.serviceit.cz/> - IT ENCYKLOPEDIE

Permanent link:
https://www.serviceit.cz/doku.php?id=it:ai:diffusion_models

Last update: 2026/01/02 12:45

