

Univerzitet u Beogradu - Elektrotehnički fakultet
Katedra za signale i sisteme

Probabilistički Difuzioni Modeli za Uklanjanje Šuma



Kandidat
Marija Brkić 2020/0200

Mentor
prof.dr Predrag Tadić

Beograd, jul 2024.

1 Rezime rada

Naredni rad predstavlja temu *Probabilističkih Difuzionih Modela za Uklanjanje Šuma*.

Difuzioni modeli su tip generativnih modela, čiji je cilj generisanje podataka visokog kvaliteta. Difuzioni model se može predstaviti nizom vremenskih stanja koja čine Markovljev lanac, nad kojima se može definisati proces difuzije (engl. forward proces) i proces obrnute difuzije (engl. backward proces). Forward proces predstavlja iterativno zašumljivanje ulaznih podataka sa ciljem da postanu što vernija aproksimacija belog Gauss-ovog šuma, dok backward proces pokušava da obrne postupak zašumljivanja i primenom varijacionog zaključivanja dođe do aprosteriorne raspodele ulaznih podataka.

Za procenu parametara gore navedene aposteriorne raspodele, u ovom radu se predlaže primena U-Net arhitekture i detaljno se objašnjava teorijska osnova difuzionih modela, kao i sam proces treniranja i njegovi rezultati.

Sadržaj

1	Rezime rada	1
2	Uvod	3
3	Struktura difuzionih modela	3
3.1	Gradivni delovi modela	3
4	Raspodele forward prelaza	5
4.1	Eksperiment	5
4.2	Parametri $\sqrt{\alpha_t}$ i $1 - \alpha_t$	7
4.3	Raspodela $q_{\Phi}(x_t x_0)$	8
4.4	Rad nad realnim podacima	10
5	Backward proces difuzionog modela	11
5.1	Donja granica dokaza	11
5.2	Donja granica dokaza u difuzionom modelu	12
5.3	Konzistentnost	14
5.4	Proračun $q_{\Phi}(x_{t-1} x_t, x_0)$	16
5.5	Trening i zaključivanje	18
5.6	Predviđanje suma	22
6	Arhitektura mreže	24
7	Rezultati i diskusija	25
7.1	Unapređenja	26
8	Zaključak	28
9	Literatura	29

2 Uvod

Difuzionni modeli su inspirisani termodinamičkim sistemima van ravnoteže [1]. U termodinamičkim sistemima van ravnoteže dolazi do procesa difuzije, pri kojem molekuli prelaze sa mesta manje koncentracije na mesto veće koncentracije, radi postizanja ravnoteže. Sa pogledom na takve procese formiran je pojam Difuzionih Probabilističkih Modela za Uklanjanje Šuma.

Difuzionni modeli se mogu predstaviti Markovljevim lancem difuzionih koraka nad kojima mogu biti definisani *forward* i *backward* procesi. U forward procesu se inkrementalno dodaje šum na podatke, dok se u backward procesu obučava model koji bi obrnuo postupak difuzije, kako bi se generisali željeni podaci iz šuma. Difuzionni modeli pripadaju grupi generativnih modela, čiji je cilj generisanje podataka visokog kvaliteta [1,2].

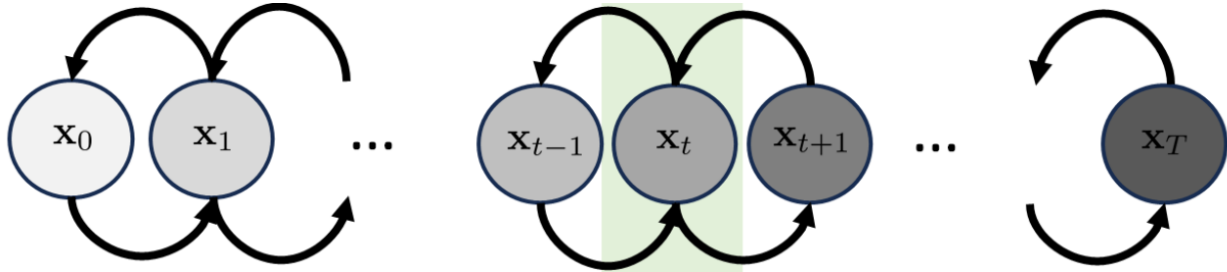
3 Struktura difuzionih modela

Struktura difuzionih modela se zove *Varijacioni Difuzioni Model*, koji sadrži sekvencu stanja x_0, x_1, \dots, x_T , pri čemu model može biti predstavljen Markovljevim procesom, što znači da stanje u trenutku t zavisi isključivo od stanja u trenutku $t - 1$:

$$p(x_t|x_{0:t-1}) = p(x_t|x_{t-1}) \quad (1)$$

1. x_0 je početno stanje koje se odnosi na početni set podataka (ulaznu sliku).
2. x_T je skrivena promenljiva, za koju bi bilo poželjno da ima raspodelu $x_T \sim N(0, I)$.
3. x_1, x_2, \dots, x_{T-1} su takođe skrivene promenljive, koje nisu beli Gauss-ov šum.

Struktura varijacionog difuzionog modela se može videti na sledećoj slici [2]:



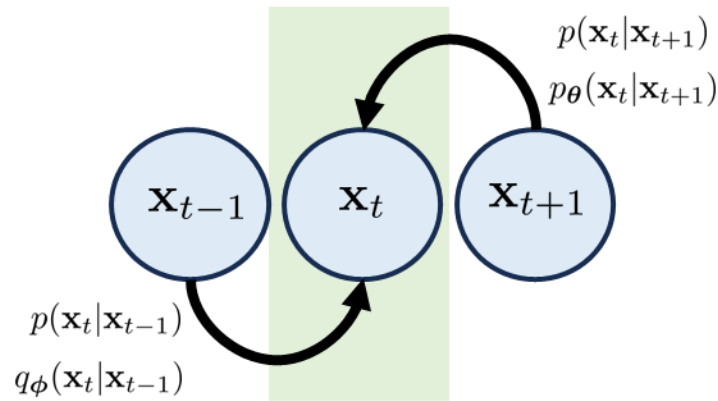
Slika 1: Struktura varijacionog difuzionog modela (preuzeto iz [2])

3.1 Gradivni delovi modela

Model prelaza u prelaznom bloku: prelazni blok sa centrom u trenutku t se sastoji od tri stanja x_{t-1} , x_t i x_{t+1} . Posmatrajući strukturu modela se može zaključiti da se do stanja x_t dolazi na dva načina: iz stanja x_{t-1} (prelaz unapred/forward proces) i iz stanja x_{t+1} (prelaz unazad/backward proces).

- Prelaz unapred se odvija iz stanja x_{t-1} u stanje x_t . Iz tog razloga je potrebno definisati model prelaza, to jest uslovnu funkciju gustine verovatnoće $p(x_t|x_{t-1})$. Ova raspodela nije poznata i zbog toga će biti aproksimirana Gauss-ovom raspodelom $q_\Phi(x_t|x_{t-1})$.
- Prelaz unazad se odvija iz stanja x_{t+1} u stanje x_t . Ovako definisan model prelaza, to jest funkcija gustine verovatnoće $p(x_t|x_{t+1})$ je takođe nepoznata i biće aproksimirana Gauss-ovom raspodelom $p_\Theta(x_t|x_{t+1})$, s tim da će matematičko očekivanje ove raspodele biti procenjeno neuralnom mrežom o kojoj će kasnije biti reči.

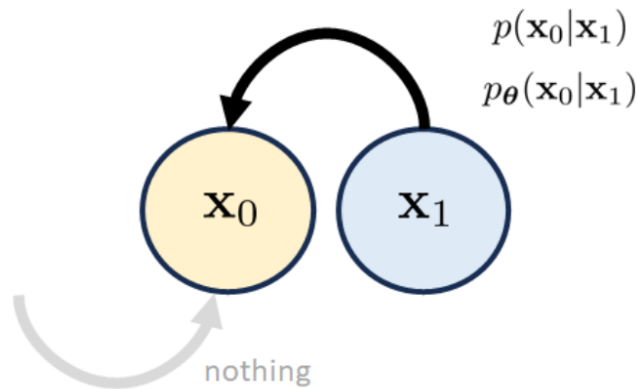
Struktura prelaznog bloka sa označenim tačnim i estimiranim modelima prelaza se može videti na sledećoj slici:



Slika 2: Struktura prelaznog gradivnog bloka (preuzeto iz [2])

Model prelaza u inicijalnom bloku: inicijalni gradivni blok u varijacionom difuzionom modelu u svom centru ima stanje x_0 . Sa obzirom na to da stanje x_{-1} ne postoji, u inicijalnom bloku može biti definisan samo model prelaza unazad, to jest uslovna funkcija gustine verovatnoće $p(x_0|x_1)$, koja je kao i u prošlom slučaju nepoznata i biće aproksimirana Gauss-ovom raspodelom $p_\Theta(x_0|x_1)$, čije će matematičko očekivanje proceniti neuralna mreža.

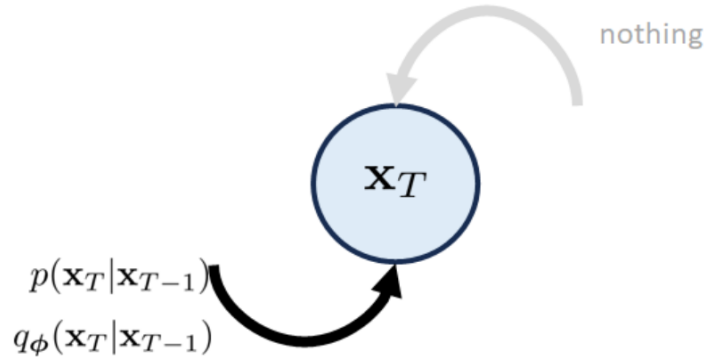
Struktura inicijalnog gradivnog bloka sa označenim tačnim i estimiranim modelom prelaza se može videti na sledećoj slici:



Slika 3: Struktura inicijalnog gradivnog bloka (preuzeto iz [2])

Model prelaza u poslednjem bloku: poslednji gradivni blok u varijacionom difuzionom modelu u svom centru ima stanje x_T . Kako je x_T poslednje stanje, za koga je pretpostavljeno da je beo Gauss-ov šum, stanje x_{T+1} ne postoji, pa može biti definisan samo model prelaza unapred, to jest uslovna funkcija gustine verovatnoće $p(x_T|x_{T-1})$. Ova funkcija gustine verovatnoće će biti aproksimirana Gauss-ovom raspodelom $q_\Phi(x_T|x_{T-1})$.

Struktura poslednjeg gradivnog bloka sa označenim tačnim i estimiranim modelom prelaza se može videti na sledećoj slici:



Slika 4: Struktura poslednjeg gradivnog bloka (preuzeto iz [2])

4 Raspodele forward prelaza

Kako je već navedeno, model prelaza iz stanja x_{t-1} u stanje x_t je definisan Gauss-ovom raspodelom, to jest funkcijom gustine verovatnoće $q_\Phi(x_t|x_{t-1})$, koja izgleda na sledeći način:

$$q_\Phi(x_t|x_{t-1}) = N(x_t|\sqrt{\alpha_t}x_{t-1}, (1 - \alpha_t)I) \quad (2)$$

gde je I jedinična matrica, α_t skalar kojim se kontroliše promena matematičkog očekivanja i kovarijacione matrice iz stanja u stanje, $\sqrt{\alpha_t}x_{t-1}$ matematičko očekivanje uslovne raspodele i $(1 - \alpha_t)I$ kovarijaciona matrica vektora x_t . Za slučajni vektor x_t je pretpostavljeno da su njegove komponente međusobno nekorelisane, zbog čega je kovarijaciona matrica jedinična matrica.

Ideja čitavog forward procesa je da se krene od početnog stanja x_0 (ulazne slike) i da se po gore navedenom modelu prelaza iz stanja u stanje kontrolisano zašumljuje. Takođe je pretpostavka da će raspodela poslednjeg stanja x_T biti normalna raspodela sa nultom srednjom vrednošću i jediničnom kovarijacionom matricom.

Kako bi se pokazalo da će ovakvim modelom prelaza raspodela sve više ličiti na normalnu raspodelu sa nultom srednjom vrednošću i jediničnom kovarijacionom matricom, za početak je predložen sledeći eksperiment [2]:

4.1 Eksperiment

Posmatrajmo bimodalnu Gauss-ovu raspodelu skalarne slučajne promenljive:

$$x_0 \sim p(x_0) = \pi_1 N(x_0|\mu_1, \sigma_1^2) + \pi_2 N(x_0|\mu_2, \sigma_2^2) \quad (3)$$

I dalje se primenjuje isti model prelaza:

$$p(x_t|x_{t-1}) = N(x_t|\sqrt{\alpha_t}x_{t-1}, (1 - \alpha_t)) \quad (4)$$

na osnovu koga se dobija:

$$x_t = \sqrt{\alpha_t}x_{t-1} + \sqrt{1 - \alpha_t}\epsilon \quad (5)$$

gde je $\epsilon \sim N(0, 1)$.

Ako bi se primenila teorema totalne verovatnoće, dobija se:

$$p(x_t|x_{t-1}) = \sum_{k=1}^2 p(x_t|x_{t-1}, k)p(k|x_{t-1}) = \sum_{k=1}^2 p(x_t|x_{t-1}, k)\pi_k \quad (6)$$

gde je k redni broj modusa iz početne bimodalne raspodele. Dalje sledi:

$$E\{x_t|x_{t-1}, k\} = \sqrt{\alpha_t}E\{x_{t-1}|k\} + \sqrt{1 - \alpha_t}E\{\epsilon\} = \sqrt{\alpha_t}\mu_k \quad (7)$$

$$\text{var}\{x_t|x_{t-1}, k\} = \alpha_t \text{var}\{x_{t-1}|k\} + (1 - \alpha_t)\text{var}\{\epsilon\} = \alpha_t\sigma_k^2 + (1 - \alpha_t) \quad (8)$$

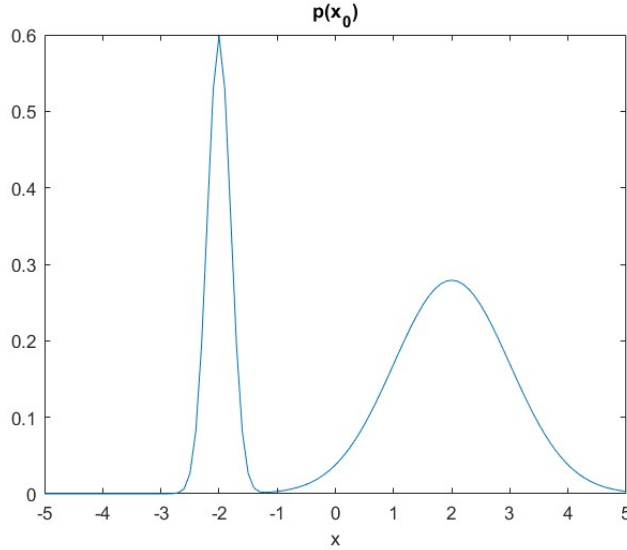
gde se podrazumeva da su x_{t-1} i ϵ međusobno nezavisni.

Odatle za $t = 1, 2, \dots, T$ sledi:

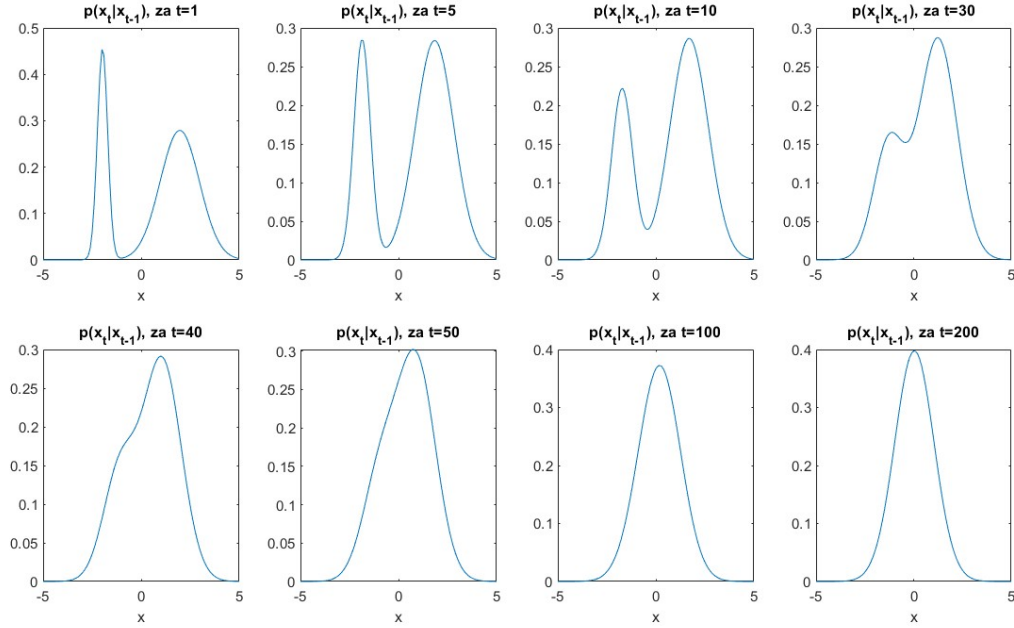
$$p(x_t|x_{t-1}) = \pi_1 N(x_t|\sqrt{\alpha_t}\mu_{1,t-1}, \alpha_t\sigma_{1,t-1}^2 + (1 - \alpha_t)) + \pi_2 N(x_t|\sqrt{\alpha_t}\mu_{2,t-1}, \alpha_t\sigma_{2,t-1}^2 + (1 - \alpha_t)) \quad (9)$$

gde je $\mu_{k,t-1}$ matematičko očekivanje u trenutku $t - 1$, dok je $\mu_{k,0} = \mu_k$. Slično je $\sigma_{k,t-1}^2$ varijansa u trenutku $t - 1$, dok je $\sigma_{k,0}^2 = \sigma_k^2$, za $k = 1, 2$.

Za uzete parametre $\pi_1 = 0.3$, $\pi_2 = 0.7$, $\mu_1 = -2$, $\mu_2 = 2$, $\sigma_1 = 0.2$, $\sigma_2 = 1$ i $\alpha_t = 0.97$ za svako t se posmatra menjanje funkcije gustine verovatnoće u vremenu. Rezultati su sledeći:



Slika 5: Početna funkcija gustine verovatnoće



Slika 6: Promena funkcija gustine verovatnoće kroz vreme

Na osnovu prethodnih grafika se jasno vidi da funkcija gustine verovatnoće posle dovoljno vremena teži Gauss-ovoj raspodeli sa nultom srednjom vrednošću i jediničnom varijansom.

4.2 Parametri $\sqrt{\alpha_t}$ i $1 - \alpha_t$

U prethodnom delu je eksperimentalno pokazano da raspodela oblika:

$$q_{\Phi}(x_t|x_{t-1}) = N(x_t|\sqrt{\alpha_t}x_{t-1}, (1 - \alpha_t)I) \quad (10)$$

za dovoljno veliko t teži Gauss-ovoj raspodeli sa nultom srednjom vrednošću i jediničnom kovarijacionom matricom, i razlog tome su upravo parametri $\sqrt{\alpha_t}$ i $1 - \alpha_t$. U cilju dokazivanja toga počnimo sa drugačijim oblikom raspodele:

$$q_{\Phi}(x_t|x_{t-1}) = N(x_t|ax_{t-1}, b^2I) \quad (11)$$

gde $a \in R$, i $b \in R$. Model prelaza se može zapisati na drugačiji način:

$$x_t = ax_{t-1} + b\epsilon_{t-1} \quad (12)$$

gde je $\epsilon_{t-1} \sim N(0, I)$.

Rekurziju iz poslednje jednakosti možemo raspisati na sledeći način:

$$x_t = ax_{t-1} + b\epsilon_{t-1} \quad (13)$$

$$= a(ax_{t-2} + b\epsilon_{t-2}) + b\epsilon_{t-1} \quad (14)$$

$$= a^2x_{t-2} + ab\epsilon_{t-2} + b\epsilon_{t-1} \quad (15)$$

$$\cdot \quad (16)$$

$$\cdot \quad (17)$$

$$\cdot \quad (18)$$

$$= a^t x_0 + b(\epsilon_{t-1} + a\epsilon_{t-2} + a^2\epsilon_{t-3} + \dots + a^{t-1}\epsilon_0) \quad (19)$$

Uzima se da je:

$$b(\epsilon_{t-1} + a\epsilon_{t-2} + a^2\epsilon_{t-3} + \dots + a^{t-1}\epsilon_0) = w_t \quad (20)$$

w_t je suma nezavisnih slučajnih promenljivih sa Gauss-ovom raspodelom. Matematičko očekivanje ove sume $E\{w_t\}$ je nula jer je matematičko očekivanje svake promenljive iz sume takođe nula. Kovarijaciona matrica se može računati na sledeći način:

$$\Sigma\{w_t\} = E\{w_t w_t^T\} = b^2(\Sigma\{\epsilon_{t-1}\} + a^2\Sigma\{\epsilon_{t-2}\} + a^4\Sigma\{\epsilon_{t-3}\} + \dots + a^{2(t-1)}\Sigma\{\epsilon_0\}) \quad (21)$$

$$= b^2(1 + a^2 + a^4 + \dots + a^{2(t-1)})I \quad (22)$$

$$= b^2 \frac{1 - a^{2t}}{1 - a^2} I \quad (23)$$

Za $0 < a < 1$ i $t \rightarrow \infty$, $a^t \rightarrow 0$. Odatle sledi:

$$\lim_{t \rightarrow \infty} \Sigma\{w_t\} = \frac{b^2}{1 - a^2} I \quad (24)$$

Kako bi $\lim_{t \rightarrow \infty} \Sigma\{w_t\} = I$, to jest da raspodela x_t teži $N(0, I)$, uzima se $a = \sqrt{\alpha}$, odakle sledi da je $b = \sqrt{1 - \alpha}$. Tada bi kovarijaciona matrica vektora x_t težila jediničnoj matrici, dok bi matematičko očekivanje ($a^t E\{x_0\}$) težilo nuli, za t koje teži beskonačnosti. Konačno ovo daje model prelaza:

$$x_t = \sqrt{\alpha}x_{t-1} + \sqrt{1 - \alpha}\epsilon_{t-1} \quad (25)$$

i uslovnu funkciju gustine verovatnoće:

$$q_\Phi(x_t|x_{t-1}) = N(x_t|\sqrt{\alpha}x_{t-1}, (1 - \alpha)I) \quad (26)$$

gde α može biti zamenjeno sa α_t .

4.3 Raspodela $q_\Phi(x_t|x_0)$

Uslovna raspodela $q_\Phi(x_t|x_0)$ može biti zabeležena i na sledeći način:

$$q_\Phi(x_t|x_0) = N(x_t|\sqrt{\bar{\alpha}_t}x_0, (1 - \bar{\alpha}_t)I) \quad (27)$$

gde je $\bar{\alpha}_t = \prod_{i=1}^t \alpha_i$.

Do te relacije se dolazi na sledeći način:

Za početak se model prelaza $q_\Phi(x_t|x_0) = N(x_t|\sqrt{\alpha_t}x_{t-1}, (1 - \alpha_t)I)$ predstavi na sledeći način:

$$x_t = \sqrt{\alpha_t}x_{t-1} + \sqrt{(1 - \alpha_t)}\epsilon_{t-1} \quad (28)$$

$$= \sqrt{\alpha_t}(\sqrt{\alpha_{t-1}}x_{t-2} + \sqrt{1 - \alpha_{t-1}}\epsilon_{t-2}) + \sqrt{1 - \alpha_t}\epsilon_{t-1} \quad (29)$$

$$\sqrt{\alpha_t\alpha_{t-1}}x_{t-2} + \sqrt{\alpha_t}\sqrt{1 - \alpha_{t-1}}\epsilon_{t-2} + \sqrt{1 - \alpha_t}\epsilon_{t-1} \quad (30)$$

gde je:

$$w_1 = \sqrt{\alpha_t}\sqrt{1 - \alpha_{t-1}}\epsilon_{t-2} + \sqrt{1 - \alpha_t}\epsilon_{t-1} \quad (31)$$

Kako je w_1 suma dva slučajna vektora sa Gauss-ovom raspodelom, i w_1 ima Gauss-ovu raspodelu, čije je matematičko očekivanje nula vektor, jer i α_{t-1} i ϵ_{t-2} imaju nulto matematičko očekivanje, dok je kovarijaciona matrica:

$$E\{w_1 w_1^T\} = ((\sqrt{\alpha_t}\sqrt{1 - \alpha_{t-1}})^2 + (\sqrt{1 - \alpha_t})^2)I \quad (32)$$

$$= (\alpha_t(1 - \alpha_{t-1}) + 1 - \alpha_t)I \quad (33)$$

$$= (1 - \alpha_t\alpha_{t-1})I \quad (34)$$

Sada jednačina (30) postaje:

$$x_t = \sqrt{\alpha_t \alpha_{t-1}} x_{t-2} + \sqrt{1 - \alpha_t \alpha_{t-1}} \epsilon_{t-2} \quad (35)$$

$$= \sqrt{\alpha_t \alpha_{t-1} \alpha_{t-2}} x_{t-3} + \sqrt{1 - \alpha_t \alpha_{t-1} \alpha_{t-2}} \epsilon_{t-3} \quad (36)$$

$$\cdot \quad (37)$$

$$\cdot \quad (38)$$

$$\cdot \quad (39)$$

$$\sqrt{\prod_{i=1}^t \alpha_i} x_0 + \sqrt{1 - \prod_{i=1}^t \alpha_i} \epsilon_0 \quad (40)$$

Definiše se $\bar{\alpha}_t = \prod_{i=1}^t \alpha_i$. Odatle sledi:

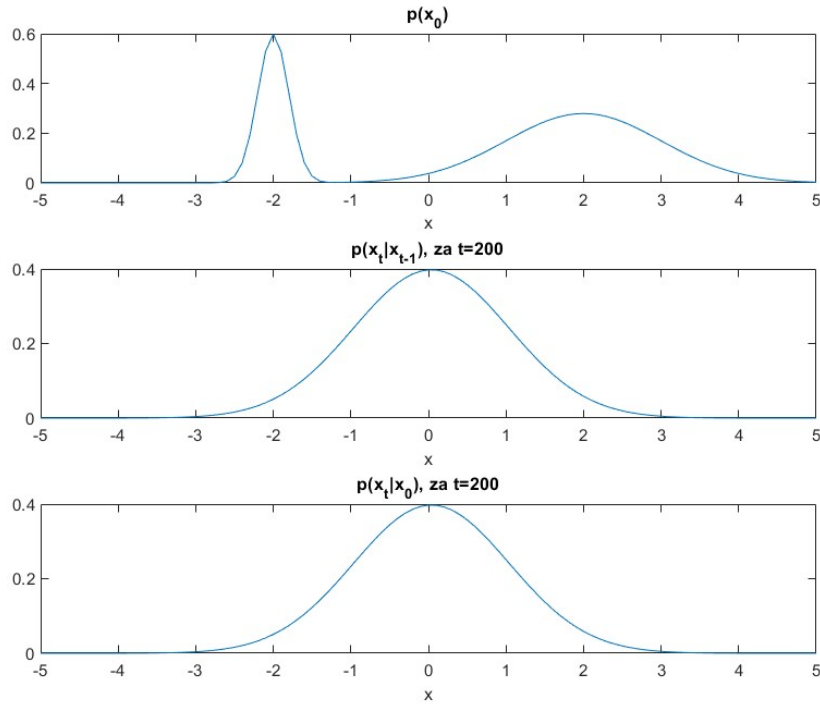
$$x_t = \sqrt{\bar{\alpha}_t} x_0 + \sqrt{1 - \bar{\alpha}_t} \epsilon_0 \quad (41)$$

Iz tog razloga je funkcija gustine verovatnoće $q_\Phi(x_t|x_0)$ definisana na sledeći način:

$$q_\Phi(x_t|x_0) = N(x_t|\sqrt{\bar{\alpha}_t}x_0, (1 - \bar{\alpha}_t)I) \quad (42)$$

Zbog ovakve definicije funkcije gustine verovatnoće za svaki trenutak t , za generisanje odbiraka slučajnog vektora x_t više nije potrebno prolaziti kroz čitav iterativan postupak, već se samo na osnovu ulaznih podataka (slike) x_0 mogu generisati odbirci vektora x_t .

Tačnost nove definicije uslovne funkcije gustine verovatnoće $q_\Phi(x_t|x_0)$ će biti eksperimentalno demonstrirana na ranijem primeru:



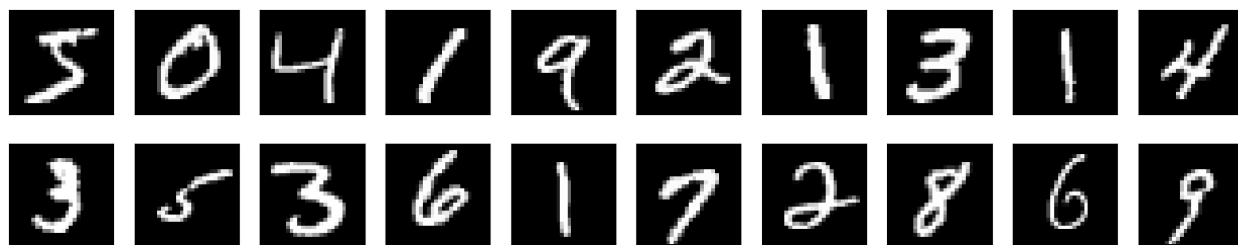
Slika 7: Eksperiment

Kao što se može videti na prethodnom grafiku, funkcija gustine verovatnoće u trenutku $t = 200$ u oba slučaja izgleda na isti način, čime je i eksperimentalno dokazan prethodan izraz.

Ovim je definisan čitav forward proces difuzionog modela, koji se svodi na polazak od ulaznih podataka (ulazne slike) na koje se iterativno dodaje šum, dok se ne dođe do seta podataka što približnijih belom Gauss-ovom šumu. Takođe je pokazano da se ovaj postupak ni ne mora raditi iterativno već se dodavanje šuma do trenutka t može odraditi direkto iz ulaznih podataka. U narednom primeru će ovaj postupak biti prikazan nad realnim podacima-slikama.

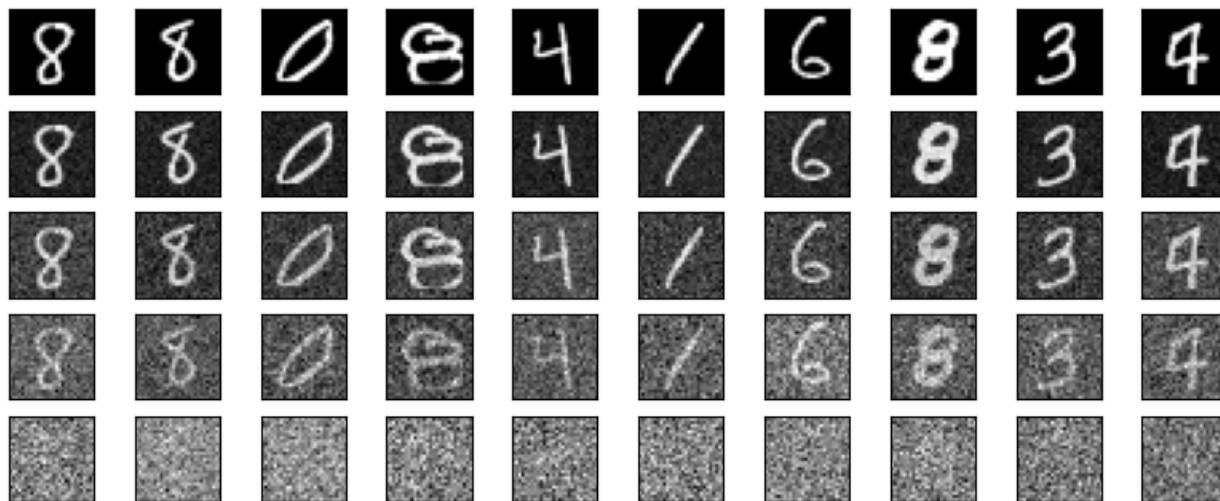
4.4 Rad nad realnim podacima

U narednom poglavlju će gore opisani postupak biti primenjen nad MNIST setom podataka koji sadrži 80000 slika rukom napisanih cifara. U okviru seta podataka slike su podeljene na trening i test skup, gde se u trening skupu nalazi 60000 slika, dok se u test skupu nalazi ostalih 20000. Slike iz korišćenog seta podataka imaju dimenzije 28x28. Neki od primera slika iz seta podataka izgledaju na sledeći način:



Slika 8: Slike iz MNIST seta podataka

Nakon učitavanja baze podataka na slike je na gore opisan način dodat šum, što izgleda na sledeći način:



Slika 9: Zašumljene slike

Na prethodnim slikama je primenjeno dodavanje šuma sa parametrom $\beta_t = 1 - \alpha_t$ koj linearno raste od vrednosti 0.0001 do 0.02 [3], za t koje uzima vrednosti [1, 300]. U prvom redu su prikazane originalne slike dok su u ostalim respektivno prikazane zašumljene slike u trenucima $t = 20, 50, 100, 300$. Na ovaj način se može primetiti da zašumljene slike zaista postaju približne belom Gauss-ovom šumu, i time je završen čitav forward proces difuzionog modela.

5 Backward proces difuzionog modela

5.1 Donja granica dokaza

Donja granica dokaza je termin koji se koristi u varijacionom zaključivanju i odnosi se na kvalitet procene aposteriorne funkcije gustine verovatnoće.

Naime, neka je nepoznata uslovna funkcija gustine verovatnoće:

$$p(Z|X) \quad (43)$$

koja se na osnovu Bayes-ove teoreme može predstaviti na sledeći način:

$$p(Z|X) = \frac{p(X|Z)p(Z)}{p(X)} \quad (44)$$

Najveći problem u zaključivanju predstavlja pronalaženje marginalne raspodele $p(X)$. Varijaciono zaključivanje predlaže alternativu pronalaženju uslovne raspodele, kojom se potpuno izbegava računanje raspodele $p(X)$. Naime, ideja je da se raspodela $p(Z|X)$ aproksimira novom raspodelom $q(Z)$ koja je jednostavnija i poznata.

Kako bi raspodela $q(Z)$ što bolje aproksimirala raspodelu $p(Z|X)$, definiše se parametar koji će govoriti o meri sličnosti te dve raspodele i naziva se Kullback-Leibler divergencija.

Kullback-Leibler divergencija je definisana na sledeći način [4]:

$$D_{KL}(q(Z)||p(Z|X)) = E_{q(Z)}\{\log \frac{q(Z)}{p(Z|X)}\} \quad (45)$$

Ukoliko bi se pretpostavilo da je Z skalarna slučajna promenljiva, Kullback-Leibler divergencija se dalje može napisati:

$$D_{KL}(q(Z)||p(Z|X)) = E_{q(Z)}\{\log \frac{q(Z)}{p(Z|X)}\} \quad (46)$$

$$= \int_z q(Z) \log \frac{q(Z)}{p(Z|X)} dz \quad (47)$$

Jednakost (47) se primenom definicije uslovne verovatnoće može predstaviti na sledeći način:

$$\int_z q(Z) \log \frac{q(Z)}{p(Z|X)} dz = \int_z q(Z) \log \frac{q(Z)p(X)}{p(Z, X)} dz \quad (48)$$

$$= \int_z q(Z) \log \frac{q(Z)}{p(Z, X)} dz + \int_z q(Z) \log(p(X)) dz \quad (49)$$

$$= E_{q(Z)}\{\log \frac{q(Z)}{p(Z, X)}\} + E_{q(Z)}\{\log(p(X))\} \quad (50)$$

$$= -E_{q(Z)}\{\log \frac{p(Z, X)}{q(Z)}\} + E_{q(Z)}\{\log(p(X))\} \quad (51)$$

$$= -E_{q(Z)}\{\log \frac{p(Z, X)}{q(Z)}\} + \log(p(X)) \quad (52)$$

Do jednakosti (52) se dolazi jer $p(X)$ ne zavisi od Z , pa može izaći ispred očekivanja, čija vrednost postaje 1.

Za logaritam dokazne promenljive $\log(p(X))$ se uvodi termin dokaz.

Aproksimacija željene uslovne raspodele se bira kao ona koja minimizira Kullback-Leibler divergenciju:

$$q^*(Z) = \arg \min_{q(Z) \in Q} D_{KL}(q(Z)||p(Z|X)) \quad (53)$$

gde je Q unapred usvojen skup raspodela $q(Z)$, koje su tipično normalne raspodele.

Takođe se definiše:

$$L(q) = E_{q(Z)} \left\{ \log \frac{p(Z, X)}{q(Z)} \right\} \quad (54)$$

Dalje se postavlja uslov da Kullback-Leibler divergencija mora biti nenegativna iz čega sledi:

$$L(q) \leq \log(p(X)) \quad (55)$$

zbog čega se $L(q)$ naziva Donja Granica Dokaza.

Sada se:

$$q^*(Z) = \arg \min_{q(Z) \in Q} D_{KL}(q(Z) || p(Z|X)) \quad (56)$$

može predstaviti kao:

$$q^*(Z) = \arg \max_{q(Z) \in Q} L(q) \quad (57)$$

gde je:

$$L(q) = E_{q(Z)} \left\{ \log \frac{p(Z, X)}{q(Z)} \right\} = -D_{KL}(q(Z) || p(Z|X)) + \log(p(X)) \quad (58)$$

5.2 Donja granica dokaza u difuzionom modelu

Donja granica dokaza u difuzionom modelu se odnosi na kvalitet procene backward uslovnih funkcija gustine verovatnoće i definisana je na sledeći način:

$$\begin{aligned} \text{ELBO}_{\Phi, \Theta}(x) = E_{q_{\Phi}(x_1|x_0)} & \left[\log \underbrace{\frac{p_{\Theta}(x_0|x_1)}{}}_{\text{kvalitet procene inicijalnog bloka}} \right] \\ & - E_{q_{\Phi}(x_{T-1}|x_0)} \left[\underbrace{D_{KL}\left(q_{\Phi}(x_T|x_{T-1}) || p(x_T)\right)}_{\text{kvalitet procene poslednjeg bloka}} \right] \\ & - \sum_{t=1}^{T-1} E_{q_{\Phi}(x_{t-1}, x_{t+1}|x_0)} \left[\underbrace{D_{KL}\left(q_{\Phi}(x_t|x_{t-1}) || p_{\Theta}(x_t|x_{t+1})\right)}_{\text{kvalitet procene prelaznog bloka}} \right]. \end{aligned} \quad (59)$$

U ovom slučaju bi moglo da se objasni intuitivno značenje svake od tri komponente donje granice dokaza difuzionog modela.

Prva komponenta se naziva rekonstrukcija i odnosi se na kvalitet početnog bloka, to jest na to koliko bi dobro početna slika x_0 mogla da se rekonstruiše iz skrivene promenljive x_1 , zbog čega u prvom članu figuriše logaritam verodostojnosti $p_{\Theta}(x_0|x_1)$.

Druga komponenta se naziva prethodno podudaranje i odnosi se na poslednji gradivni blok difuzionog modela. U drugoj komponenti figuriše Kullback-Leibler (KL) divergencija između procene funkcije gustine verovatnoće $q_{\Phi}(x_T|x_{T-1})$ koja je dobijena forward procesom, i željene funkcije gustine verovatnoće u krajnjem bloku $p(x_T)$ koja je $N(0, I)$.

Treća komponenta se naziva konzistentnost i odnosi se na kvalitet procene svih prelaznih blokova. Naime, kao što je već objašnjeno, do odбирaka x_t se može doći na dva načina, preko forward raspodele $q_{\Phi}(x_t|x_{t-1})$ i backward raspodele $p_{\Theta}(x_t|x_{t+1})$, stoga u ovoj komponenti figuriše KL divergencija između te dve raspodele.

Na ovaj način se donjom granicom dokaza obuhvata procena kvaliteta čitavog modela.

Nadalje je opisan način dolaska do izraza (59). Za početak se definiše sledeća notacija:

$$x_{0:T} = \{x_0, x_1, \dots, x_T\} \quad (60)$$

U ovako definisanom problemu pojam dokaz se odnosi na logaritam dokazne promenljive, što je ulazna slika:

$$\log(p(x)) \quad (61)$$

$$= \log(p(x_0)) \quad (62)$$

$$= \log \int p(x_{0:T}) dx_{1:T} \quad (63)$$

$$= \log \int p(x_{0:T}) \frac{q_{\Phi}(x_{1:T}|x_0)}{q_{\Phi}(x_{1:T}|x_0)} dx_{1:T} \quad (64)$$

$$= \log \int q_{\Phi}(x_{1:T}|x_0) \frac{p(x_{0:T})}{q_{\Phi}(x_{1:T}|x_0)} dx_{1:T} \quad (65)$$

$$= \log E_{q_{\Phi}(x_{1:T}|x_0)} \left\{ \frac{p(x_{0:T})}{q_{\Phi}(x_{1:T}|x_0)} \right\} \quad (66)$$

Nadalje se primenjuje Jansen-ova nejednakost koja tvrdi da za bilo koju slučajnu promenljivu X i konkavnu funkciju f važi da:

$$f(E\{X\}) \geq E\{f(X)\} \quad (67)$$

Kako je logaritamska funkcija konkavna funkcija, dalje sledi:

$$\log(x) \quad (68)$$

$$= \log E_{q_{\Phi}(x_{1:T}|x_0)} \left\{ \frac{p(x_{0:T})}{q_{\Phi}(x_{1:T}|x_0)} \right\} \quad (69)$$

$$\geq E_{q_{\Phi}(x_{1:T}|x_0)} \left\{ \log \frac{p(x_{0:T})}{q_{\Phi}(x_{1:T}|x_0)} \right\} \quad (70)$$

Na osnovu grafa koji predstavlja sekvencu stanja difuzionog modela se združena raspodela $p(x_{0:T})$ može predstaviti lančanim pravilom na sledeći način:

$$p(x_{0:T}) = p(x_T) \prod_{t=1}^T p(x_{t-1}|x_t) = p(x_T)p(x_0|x_1) \prod_{t=2}^T p(x_{t-1}|x_t) \quad (71)$$

Na sličan način se može doći i do raspodele $q_{\Phi}(x_{1:T}|x_0)$:

$$q_{\Phi}(x_{1:T}|x_0) = \prod_{t=1}^T q_{\Phi}(x_t|x_{t-1}) = q_{\Phi}(x_T|x_{T-1}) \prod_{t=1}^{T-1} q_{\Phi}(x_t|x_{t-1}) \quad (72)$$

Na osnovu jednačina (70), (71) i (72) se dobija:

$$\log(x) \quad (73)$$

$$\geq E_{q_{\Phi}(x_{1:T}|x_0)} \left\{ \log \frac{p(x_{0:T})}{q_{\Phi}(x_{1:T}|x_0)} \right\} \quad (74)$$

$$= E_{q_{\Phi}(x_{1:T}|x_0)} \left\{ \log \frac{p(x_T)p(x_0|x_1) \prod_{t=2}^T p(x_{t-1}|x_t)}{q_{\Phi}(x_T|x_{T-1}) \prod_{t=1}^{T-1} q_{\Phi}(x_t|x_{t-1})} \right\} \quad (75)$$

$$= E_{q_{\Phi}(x_{1:T}|x_0)} \left\{ \log \frac{p(x_T)p(x_0|x_1) \prod_{t=1}^{T-1} p(x_t|x_{t+1})}{q_{\Phi}(x_T|x_{T-1}) \prod_{t=1}^{T-1} q_{\Phi}(x_t|x_{t-1})} \right\} \quad (76)$$

$$= E_{q_{\Phi}(x_{1:T}|x_0)} \left\{ \log \frac{p(x_T)p(x_0|x_1)}{q_{\Phi}(x_T|x_{T-1})} \right\} + E_{q_{\Phi}(x_{1:T}|x_0)} \left\{ \log \frac{\prod_{t=1}^{T-1} p(x_t|x_{t+1})}{\prod_{t=1}^{T-1} q_{\Phi}(x_t|x_{t-1})} \right\} \quad (77)$$

$$(78)$$

Prvi od prethodna dva sabirka se može napisati na sledeći način:

$$E_{q_{\Phi}(x_{1:T}|x_0)}\left\{\log \frac{p(x_T)p(x_0|x_1)}{q_{\Phi}(x_T|x_{T-1})}\right\} \quad (79)$$

$$= E_{q_{\Phi}(x_{1:T}|x_0)}\{\log(p(x_0|x_1))\} + E_{q_{\Phi}(x_{1:T}|x_0)}\left\{\log \frac{p(x_T)}{q_{\Phi}(x_T|x_{T-1})}\right\} \quad (80)$$

gde se prvi sabirak odnosi na rekonstrukciju, dok se drugi sabirak odnosi na prethodno podudaranje.

Rekonstrukcija se može pojednostaviti na sledeći način:

$$E_{q_{\Phi}(x_{1:T}|x_0)}\{\log(p(x_0|x_1))\} = E_{q_{\Phi}(x_1|x_0)}\{\log(p(x_0|x_1))\} \quad (81)$$

Prethodno podudaranje se može predstaviti na sledeći način:

$$E_{q_{\Phi}(x_{1:T}|x_0)}\left\{\log \frac{p(x_T)}{q_{\Phi}(x_T|x_{T-1})}\right\} \quad (82)$$

$$= E_{q_{\Phi}(x_T, x_{T-1}|x_0)}\left\{\log \frac{p(x_T)}{q_{\Phi}(x_T|x_{T-1})}\right\} \quad (83)$$

$$= -E_{q_{\Phi}(x_{T-1}|x_0)}\{D_{KL}(q_{\Phi}(x_T|x_{T-1})||p(x_T))\} \quad (84)$$

gde se iz oblika (82) u oblik (83) prelazi jer $\log \frac{p(x_T)}{q_{\Phi}(x_T|x_{T-1})}$ zavisi samo od x_T i x_{T-1} .

Konačno, drugi sabirak iz jednakosti (77) se sređuje na sledeći način:

$$E_{q_{\Phi}(x_{1:T}|x_0)}\left\{\log \frac{\prod_{t=1}^{T-1} p(x_t|x_{t+1})}{\prod_{t=1}^{T-1} q_{\Phi}(x_t|x_{t-1})}\right\} \quad (85)$$

$$= E_{q_{\Phi}(x_{1:T}|x_0)}\left\{\log \prod_{t=1}^{T-1} \frac{p(x_t|x_{t+1})}{q_{\Phi}(x_t|x_{t-1})}\right\} \quad (86)$$

$$= \sum_{t=1}^{T-1} E_{q_{\Phi}(x_{1:T}|x_0)}\left\{\log \frac{p(x_t|x_{t+1})}{q_{\Phi}(x_t|x_{t-1})}\right\} \quad (87)$$

$$= \sum_{t=1}^{T-1} E_{q_{\Phi}(x_{t-1}, x_t, x_{t+1}|x_0)}\left\{\log \frac{p(x_t|x_{t+1})}{q_{\Phi}(x_t|x_{t-1})}\right\} \quad (88)$$

$$= - \sum_{t=1}^{T-1} E_{q_{\Phi}(x_{t-1}, x_{t+1}|x_0)}\{D_{KL}(q_{\Phi}(x_t|x_{t-1})||p(x_t|x_{t+1}))\} \quad (89)$$

čime se dolazi do konzistentnosti. Zamenom $p(x_t|x_{t+1})$ sa $p_{\Theta}(x_t|x_{t+1})$ i $p(x_0|x_1)$ sa $p_{\Theta}(x_0|x_1)$ se dolazi do oblika za donju granicu dokaza iz jednakosti (59).

5.3 Konzistentnost

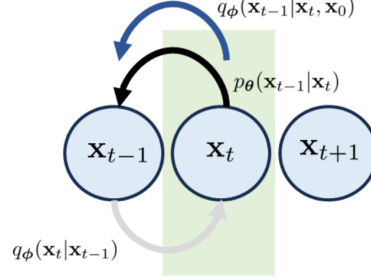
Posmatrajući element konzistentnosti primećuje se da se generisanje raspodela $p_{\Theta}(x_t|x_{t+1})$ i $q_{\Phi}(x_t|x_{t-1})$ kreće u suprotnim smerovima i za uzorkovanje svakog x_t su potrebni odbirci i x_{t-1} i x_{t+1} , što nije lako izvodljivo. Zbog toga se pristupa modifikaciji koja koristi Bayes-ovu teorem:

$$q(x_t|x_{t-1}) = \frac{q(x_{t-1}|x_t)q(x_t)}{q(x_{t-1})} \quad (90)$$

Kada se doda uslovljenost sa x_0 prethodna jednačina izgleda na sledeći način:

$$q(x_t|x_{t-1}, x_0) = \frac{q(x_{t-1}|x_t, x_0)q(x_t|x_0)}{q(x_{t-1}|x_0)} \quad (91)$$

Na ovaj jednostavan način se raspodela $q(x_t|x_{t-1}, x_0)$ zapisuje preko raspodele $q(x_{t-1}|x_t, x_0)$, čime se postiže da se proračun raspodela $q_\Phi(x_{t-1}|x_t, x_0)$ i $p_\Theta(x_t|x_{t+1})$ vrši u istom smeru. Reprezentacija ovoga se može videti na sledećoj slici:



Slika 10: Prelazak na $q_\Phi(x_t|x_{t-1}, x_0)$ (preuzeto iz [2])

Sada će u konzistentnosti figurisati KL divergencija između raspodela $q_\Phi(x_{t-1}|x_t, x_0)$ i $p_\Theta(x_t|x_{t+1})$. Prethodno objašnjenom modifikacijom se dolazi do sledećeg oblika donje granice dokaza:

$$\begin{aligned}
 \text{ELBO}_{\Phi, \Theta}(x) &= E_{q_\Phi(x_1|x_0)} \left[\log \underbrace{p_\Theta(x_0|x_1)}_{\text{kvalitet procene inicijalnog bloka}} \right] \\
 &\quad - \underbrace{D_{\text{KL}}(q_\Phi(x_T|x_0) \| p(x_T))}_{\text{novi kvalitet procene poslednjeg bloka}} \\
 &\quad - \sum_{t=2}^T E_{q_\Phi(x_t|x_0)} \left[\underbrace{D_{\text{KL}}(q_\Phi(x_{t-1}|x_t, x_0) \| p_\Theta(x_{t-1}|x_t))}_{\text{novi kvalitet procene prelaznog bloka}} \right]. \tag{92}
 \end{aligned}$$

Donja granica dokaza se ponovo sastoji iz tri komponente.

Prva komponenta, to jest komponenta rekonstrukcije je zadržala isti oblik.

Druga komponenta, to jest prethodno podudaranje je iz prethodnog oblika $-E_{q_\Phi(x_{T-1}|x_0)} \{D_{\text{KL}}(q_\Phi(x_T|X_{T-1}) \| p(x_T))\}$ prešlo u oblik $-D_{\text{KL}}(q_\Phi(x_T|x_0) \| p(x_T))$ zato što više nijedan član u KL divergenciji ne zavisi od x_{T-1} , pa matematičko očekivanje postaje suvišno.

Treća komponenta, to jest konzistentnost se i dalje odnosi na kvalitet procene prelaznog bloka, sa time da su unete dve promene. Prva je da t počinje od 2 i završava se u T , dok je ranije počinjalo u 1 i završavalo se u $T-1$. Druga promena je da se sada KL divergencija traži između raspodela $q_\Phi(x_{t-1}|x_t, x_0)$ i $p_\Theta(x_{t-1}|x_t)$, i time se efektivno traže parametri raspodele q_Φ koja će aproksimirati proces obrnute difuzije.

Sa ciljem dokaza poslednje definicije donje granice dokaza difuzionog modela, vraćamo se na jednakost (70):

$$\log(x) \tag{93}$$

$$\geq E_{q_\Phi(x_{1:T}|x_0)} \left\{ \log \frac{p(x_{0:T})}{q_\Phi(x_{1:T}|x_0)} \right\} \tag{94}$$

$$= E_{q_\Phi(x_{1:T}|x_0)} \left\{ \log \frac{p(x_T)p(x_0|x_1) \prod_{t=2}^T p(x_{t-1}|x_t)}{q_\Phi(x_1|x_0) \prod_{t=2}^T q_\Phi(x_t|x_{t-1}, x_0)} \right\} \tag{95}$$

$$= E_{q_\Phi(x_{1:T}|x_0)} \left\{ \log \frac{p(x_T)p(x_0|x_1)}{q_\Phi(x_1|x_0)} \right\} + E_{q_\Phi(x_{1:T}|x_0)} \left\{ \log \frac{\prod_{t=2}^T p(x_{t-1}|x_t)}{\prod_{t=2}^T q_\Phi(x_t|x_{t-1}, x_0)} \right\} \tag{96}$$

Krenimo od proizvoda u drugom sabirku:

$$\prod_{t=2}^T \frac{p(x_{t-1}|x_t)}{q_{\Phi}(x_t|x_{t-1}, x_0)} \quad (97)$$

$$\prod_{t=2}^T \frac{p(x_{t-1}|x_t)}{\frac{q_{\Phi}(x_{t-1}|x_t, x_0)q_{\Phi}(x_t|x_0)}{q_{\Phi}(x_{t-1}|x_0)}} \quad (98)$$

$$\prod_{t=2}^T \frac{p(x_{t-1}|x_t)}{q_{\Phi}(x_{t-1}|x_t, x_0)} \prod_{t=2}^T \frac{q_{\Phi}(x_{t-1}|x_0)}{q_{\Phi}(x_t|x_0)} \quad (99)$$

$$\prod_{t=2}^T \frac{p(x_{t-1}|x_t)}{q_{\Phi}(x_{t-1}|x_t, x_0)} \frac{q_{\Phi}(x_1|x_0)}{q_{\Phi}(x_T|x_0)} \quad (100)$$

gde se do poslednje jednakosti dolazi jednostavnim skraćivanjem činilaca koji su isti i u brojiocu i u imeniocu.

Povratkom na izraz (96) dobija se:

$$= E_{q_{\Phi}(x_{1:T}|x_0)} \left\{ \log \frac{p(x_T)p(x_0|x_1)}{q_{\Phi}(x_1|x_0)} \right\} + E_{q_{\Phi}(x_{1:T}|x_0)} \left\{ \log \frac{\prod_{t=2}^T p(x_{t-1}|x_t)}{\prod_{t=2}^T q_{\Phi}(x_t|x_{t-1}, x_0)} \right\} \quad (101)$$

$$= E_{q_{\Phi}(x_{1:T}|x_0)} \left\{ \log \frac{p(x_T)p(x_0|x_1)}{q_{\Phi}(x_1|x_0)} + \log \left(\frac{q_{\Phi}(x_1|x_0)}{q_{\Phi}(x_T|x_0)} \right) \right\} + E_{q_{\Phi}(x_{1:T}|x_0)} \left\{ \log \prod_{t=2}^T \frac{p(x_{t-1}|x_t)}{q_{\Phi}(x_{t-1}|x_t, x_0)} \right\} \quad (102)$$

$$= E_{q_{\Phi}(x_{1:T}|x_0)} \left\{ \log \frac{p(x_T)p(x_0|x_1)}{q_{\Phi}(x_T|x_0)} \right\} + E_{q_{\Phi}(x_{1:T}|x_0)} \left\{ \log \prod_{t=2}^T \frac{p(x_{t-1}|x_t)}{q_{\Phi}(x_{t-1}|x_t, x_0)} \right\} \quad (103)$$

Prvi član poslednjeg zbira se može predstaviti na sledeći način:

$$E_{q_{\Phi}(x_{1:T}|x_0)} \left\{ \log \frac{p(x_T)p(x_0|x_1)}{q_{\Phi}(x_T|x_0)} \right\} \quad (104)$$

$$= E_{q_{\Phi}(x_{1:T}|x_0)} \{p(x_0|x_1)\} + E_{q_{\Phi}(x_{1:T}|x_0)} \left\{ \log \frac{p(x_T)}{q_{\Phi}(x_T|x_0)} \right\} \quad (105)$$

$$= E_{q_{\Phi}(x_1|x_0)} \{p(x_0|x_1)\} - D_{KL}(q_{\Phi}(x_T|x_0)||p(x_T)) \quad (106)$$

gde se prvi sabirak odnosi na rekonstrukciju, dok se drugi odnosi na prethodno podudaranje.

Drugi član iz jednakosti (103) se može raspisati na sledeći način:

$$E_{q_{\Phi}(x_{1:T}|x_0)} \left\{ \log \prod_{t=2}^T \frac{p(x_{t-1}|x_t)}{q_{\Phi}(x_{t-1}|x_t, x_0)} \right\} \quad (107)$$

$$= \sum_{t=2}^T E_{q_{\Phi}(x_{t-1}, x_t|x_0)} \left\{ \log \frac{p(x_{t-1}|x_t)}{q_{\Phi}(x_{t-1}|x_t, x_0)} \right\} \quad (108)$$

$$= - \sum_{t=2}^T E_{q_{\Phi}(x_t|x_0)} \{D_{KL}(q_{\Phi}(x_{t-1}|x_t, x_0)||p(x_{t-1}|x_t))\} \quad (109)$$

što je konzistentnost.

Sada se zamenom $p(x_{t-1}|x_t)$ sa $p_{\Theta}(x_{t-1}|x_t)$ i $p(x_0|x_1)$ sa $p_{\Theta}(x_0|x_1)$ dolazi do izraza (92).

5.4 Proračun $q_{\Phi}(x_{t-1}|x_t, x_0)$

Kao i dosadašnje q_{Φ} raspodele, i raspodela $q_{\Phi}(x_{t-1}|x_t, x_0)$ zadržava oblik Gauss-ove raspodele, koja je potpuno određena svojim matematičkim očekivanjem i kovarijacionom matricom i data je na sledeći način:

$$q_{\Phi}(x_{t-1}|x_t, x_0) = N(x_{t-1}|\mu_q(x_t, x_0), \Sigma_q(t)I) \quad (110)$$

gde je:

$$\mu_q(x_t, x_0) = \frac{(1 - \bar{\alpha}_{t-1})\sqrt{\alpha_t}}{1 - \bar{\alpha}_t}x_t + \frac{(1 - \alpha_t)\sqrt{\bar{\alpha}_{t-1}}}{1 - \bar{\alpha}_t}x_0 \quad (111)$$

$$\Sigma_q(t) = \frac{(1 - \alpha_t)(1 - \sqrt{\bar{\alpha}_{t-1}})}{1 - \bar{\alpha}_t}I = \sigma_q^2(t)I \quad (112)$$

Prethodna jednakost se može dokazati na sledeći način:

Počevši od Bayes-ove teoreme se dobija sledeće:

$$q_\Phi(x_{t-1}|x_t, x_0) = \frac{q_\Phi(x_t|x_{t-1}, x_0)q_\Phi(x_{t-1}|x_0)}{q_\Phi(x_t|x_0)} = \frac{N(x_t|\sqrt{\alpha_t}x_{t-1}, (1 - \alpha_t)I)N(x_{t-1}|\sqrt{\bar{\alpha}_{t-1}}x_0, (1 - \bar{\alpha}_{t-1})I)}{N(x_t|\sqrt{\bar{\alpha}_t}x_0, (1 - \bar{\alpha}_t)I)} \quad (113)$$

Radi pojednostavljena će biti posmatran skalarni slučaj:

$$q_\Phi(x_{t-1}|x_t, x_0) \propto \exp\left\{\frac{(x_t - \sqrt{\alpha_t}x_{t-1})^2}{2(1 - \alpha_t)} + \frac{(x_{t-1} - \sqrt{\bar{\alpha}_{t-1}}x_0)^2}{2(1 - \bar{\alpha}_{t-1})} - \frac{(x_t - \sqrt{\bar{\alpha}_t}x_0)^2}{2(1 - \bar{\alpha}_t)}\right\} \quad (114)$$

Dalje se uvode sledeće smene $x = x_t$, $y = x_{t-1}$, $z = x_0$, $a = \alpha_t$, $b = \bar{\alpha}_{t-1}$ i $c = \bar{\alpha}_t$.

Posmatra se kvadratna funkcija:

$$f(y) = \frac{(x - \sqrt{a}y)^2}{2(1 - a)} + \frac{(y - \sqrt{b}z)^2}{2(1 - b)} - \frac{(x - \sqrt{c}z)^2}{2(1 - c)} \quad (115)$$

Kako kvadratna funkcija ima oblik parabole, matematičko očekivanje raspodele $q_\Phi(x_{t-1}|x_t, x_0)$ se nalazi u ekstremumu parabole. Zato je potrebno naći prvi izvod kvadratne funkcije:

$$f'(y) = \frac{1 - ab}{(1 - a)(1 - b)}y - \left(\frac{\sqrt{a}}{1 - a}x + \frac{\sqrt{b}}{1 - b}z\right) \quad (116)$$

Izjednačavanjem prvog izvoda sa nulom se dobija:

$$y = \frac{(1 - b)\sqrt{a}}{1 - ab}x + \frac{(1 - a)\sqrt{b}}{1 - ab}z \quad (117)$$

Vraćanjem prethodnih smena i uviđanjem da je $\alpha_t\bar{\alpha}_{t-1} = \bar{\alpha}_t$ se dobija željeno matematičko očekivanje:

$$\mu_q(x_t, x_0) = \frac{(1 - \bar{\alpha}_{t-1})\sqrt{\alpha_t}}{1 - \bar{\alpha}_t}x_t + \frac{(1 - \alpha_t)\sqrt{\bar{\alpha}_{t-1}}}{1 - \bar{\alpha}_t}x_0 \quad (118)$$

Na sličan način, varijansa se može aproksimirati recipročnom vrednošću drugog izvoda kvadratne funkcije (115) na sledeći način:

$$f''(y) = \frac{1 - ab}{(1 - a)(1 - b)} = \frac{1 - \bar{\alpha}_t}{(1 - \alpha_t)(1 - \sqrt{\bar{\alpha}_{t-1}})} \quad (119)$$

Na taj način kovarijaciona matrica raspodele $q_\Phi(x_{t-1}|x_t, x_0)$ postaje:

$$\Sigma_q(t) = \frac{(1 - \alpha_t)(1 - \sqrt{\bar{\alpha}_{t-1}})}{1 - \bar{\alpha}_t}I = \sigma_q^2(t)I \quad (120)$$

Pozitivna stvar kod ovako definisane uslovne funkcije gustine verovatnoće $q_\Phi(x_{t-1}|x_t, x_0)$ je to što je ona potpuno zavisna od x_t i x_0 i samim tim nije potrebno da mreža procenjuje bilo koji od njenih parametara.

Kada se posmatra komponenta konzistentnosti donje granice dokaza difuzionog modela, ona predstavlja sumu očekivanja KL divergencija koje imaju sledeći oblik:

$$D_{KL}(q_\Phi(x_{t-1}|x_t, x_0)||p_\Theta(x_{t-1}|x_t)) \quad (121)$$

u kome u raspodeli $q_\Phi(x_{t-1}|x_t, x_0)$ nema parametara koji treba da se "nauče", međutim raspodela $p_\Theta(x_{t-1}|x_t)$ nije poznata i biće potrebno da mreža nauči neke njene parametre.

Za početak se takođe uvodi pretpostavka da je raspodela $p_{\Theta}(x_{t-1}|x_t)$ Gauss-ova raspodela koja će imati sledeći oblik:

$$p_{\Theta}(x_{t-1}|x_t) = N(x_{t-1}|\mu_{\Theta}(x_t), \sigma_q^2(t)I) \quad (122)$$

gde će matematičko očekivanje biti procenjeno neuralnom mrežom, dok se za kovarijacionu matricu uzima ista kao i za raspodelu $q_{\Phi}(x_{t-1}|x_t, x_0)$, što je pretpostavka koja je uvedena isključivo zbog matematičkog pojednostavljenja. Sada KL divergencija izgleda na sledeći način:

$$D_{KL}(q_{\Phi}(x_{t-1}|x_t, x_0)||p_{\Theta}(x_{t-1}|x_t)) \quad (123)$$

$$= D_{KL}(N(x_{t-1}|\mu_q(x_t, x_0), \Sigma_q(t)I)||N(x_{t-1}|\mu_{\Theta}(x_t), \sigma_q^2(t)I)) \quad (124)$$

$$= \frac{1}{2\sigma_q^2(t)} \|\mu_q(x_t, x_0) - \mu_{\Theta}(x_t)\|^2 \quad (125)$$

Poslednja jednakost je dobijena na osnovu osobine KL divergencije koja glasi da je KL divergencija između dve Gauss-ove raspodele istih kovarijacionih matrica jednaka kvadratu Euklidskog rastojanja između vektora matematičkih očekivanja te dve raspodele.

Vrativši se na donju granicu dokaza difuzionog modela, sada njen zapis može biti sledeći:

$$\begin{aligned} \text{ELBO}_{\Theta}(x) &= E_{q(x_1|x_0)} \left[\log p_{\Theta}(x_0|x_1) \right] \\ &\quad - \underbrace{D_{KL}(q(x_T|x_0)||p(x_T))}_{\text{ništa što mreža procenjuje}} \\ &\quad - \sum_{t=2}^T E_{q(x_t|x_0)} \left[\frac{1}{2\sigma_q^2(t)} \|\mu_q(x_t, x_0) - \mu_{\Theta}(x_t)\|^2 \right]. \end{aligned} \quad (126)$$

U odnosu na zapis (92) su unete sledeće izmene:

1. Svaki parametar Φ je uklonjen, jer bi to bili parametri forward procesa koje bi mreža trebalo da proceni, ali je pokazano da u tom procesu nije potrebno da se bilo šta procenjuje, već je sve jasno definisano ulaznim podacima (slikom) x_0 .
2. Parametri Θ su parametri (težine) mreže koja procenjuje matematičko očekivanje $\mu_{\Theta}(x_t)$.
3. Uzorkovanje iz raspodele $q(x_t|x_0)$ se vrši na osnovu jednakosti (42).
4. Parametar $E_{q(x_1|x_0)}\{\log p_{\Theta}(x_0|x_1)\}$ će moći da se proceni čim x_1 bude poznato i mreža može da proceni $\mu_{\Theta}(x_1)$.

5.5 Trening i zaključivanje

Poslednji zaključak izveden iz donje granice dokaza difuzionog modela je da je potrebno istrenirati mrežu koja će minimizirati sledeću kriterijumsku funkciju:

$$\frac{1}{2\sigma_q^2(t)} \|\mu_q(x_t, x_0) - \mu_{\Theta}(x_t)\|^2 \quad (127)$$

gde je $\mu_q(x_t, x_0)$ poznato, dok se $\mu_{\Theta}(x_t)$ procenjuje.

Kako je:

$$\mu_q(x_t, x_0) = \frac{(1 - \bar{\alpha}_{t-1})\sqrt{\alpha_t}}{1 - \bar{\alpha}_t} x_t + \frac{(1 - \alpha_t)\sqrt{\bar{\alpha}_{t-1}}}{1 - \bar{\alpha}_t} x_0 \quad (128)$$

i za $\mu_{\Theta}(x_t)$ će radi jednostavnosti biti uvedena slična forma:

$$\mu_{\Theta}(x_t) = \frac{(1 - \bar{\alpha}_{t-1})\sqrt{\alpha_t}}{1 - \bar{\alpha}_t} x_t + \frac{(1 - \alpha_t)\sqrt{\bar{\alpha}_{t-1}}}{1 - \bar{\alpha}_t} \hat{x}_{\Theta}(x_t) \quad (129)$$

gde je $\hat{x}_{\Theta}(x_t)$ novi parametar koji mreža treba da proceni.

Koristeći jednakosti (129) i (128) u (127) se dobija sledeće:

$$\frac{1}{2\sigma_q^2(t)} \|\mu_q(x_t, x_0) - \mu_\Theta(x_t)\|^2 \quad (130)$$

$$= \frac{1}{2\sigma_q^2(t)} \left\| \frac{(1 - \alpha_t)\sqrt{\bar{\alpha}_{t-1}}}{1 - \bar{\alpha}_t} (\hat{x}_\Theta(x_t) - x_0) \right\|^2 \quad (131)$$

$$= \frac{1}{2\sigma_q^2(t)} \frac{(1 - \alpha_t)\sqrt{\bar{\alpha}_{t-1}}}{1 - \bar{\alpha}_t} \|\hat{x}_\Theta(x_t) - x_0\|^2 \quad (132)$$

Sada donja granica dokaza difuzionog modela može biti pojednostavljena na sledeći način:

$$\begin{aligned} \text{ELBO}_\Theta(x) &= E_{q(x_1|x_0)} \left[\log p_\Theta(x_0|x_1) \right] - \sum_{t=2}^T E_{q(x_t|x_0)} \left[\frac{1}{2\sigma_q^2(t)} \|\mu_q(x_t, x_0) - \mu_\Theta(x_t)\|^2 \right] \\ &= E_{q(x_1|x_0)} \left[\log p_\Theta(x_0|x_1) \right] - \sum_{t=2}^T E_{q(x_t|x_0)} \left[\frac{1}{2\sigma_q^2(t)} \frac{(1 - \alpha_t)^2 \bar{\alpha}_{t-1}}{(1 - \bar{\alpha}_t)^2} \|\hat{x}_\Theta(x_t) - x_0\|^2 \right] \end{aligned}$$

Za prvi sabirak važi:

$$\log p_\Theta(x_0|x_1) = \log N(x_0|\mu_\Theta(x_1), \sigma_q^2(1)I) \quad (133)$$

$$\propto -\frac{1}{2\sigma_q^2(1)} \|\mu_\Theta(x_1) - x_0\|^2 \quad (134)$$

$$= -\frac{1}{2\sigma_q^2(1)} \left\| \frac{(1 - \bar{\alpha}_0)\sqrt{\bar{\alpha}_1}}{1 - \bar{\alpha}_1} x_1 + \frac{(1 - \alpha_1)\sqrt{\bar{\alpha}_0}}{1 - \bar{\alpha}_1} \hat{x}_\Theta(x_1) - x_0 \right\|^2 \quad (135)$$

$$= -\frac{1}{2\sigma_q^2(1)} \left\| \frac{(1 - \alpha_1)}{1 - \bar{\alpha}_1} \hat{x}_\Theta(x_1) - x_0 \right\|^2 \quad (136)$$

$$= -\frac{1}{2\sigma_q^2(1)} \left\| \hat{x}_\Theta(x_1) - x_0 \right\|^2 \quad (137)$$

gde se iz koraka (135) u korak (136) prelazi jer je $\alpha_0 = 1$, a iz koraka (136) u korak (137) jer je $\bar{\alpha}_1 = \alpha_1$.

Konačno se za donju granicu dokaza difuzionog modela dobija:

$$\text{ELBO}_\Theta(x) = -\sum_{t=1}^T E_{q(x_t|x_0)} \left[\frac{1}{2\sigma_q^2(t)} \frac{(1 - \alpha_t)^2 \bar{\alpha}_{t-1}}{(1 - \bar{\alpha}_t)^2} \|\hat{x}_\Theta(x_t) - x_0\|^2 \right] \quad (138)$$

Pošto je cilj modela da maksimizuje donju granicu dokaza, kriterijumska funkcija(funkcija gubitka) difuzionog probabilističkog modela se svodi na izraz:

$$\Theta^* = \arg \min_{\Theta} \sum_{t=1}^T \frac{1}{2\sigma_q^2(t)} \frac{(1 - \alpha_t)^2 \bar{\alpha}_{t-1}}{(1 - \bar{\alpha}_t)^2} E_{q(x_t|x_0)} \left[\|\hat{x}_\Theta(x_t) - x_0\|^2 \right] \quad (139)$$

Ako bi se zanemarile konstante i očekivanje dobija se:

$$\Theta^* = \|\hat{x}_\Theta(x_t) - x_0\|^2 \quad (140)$$

Ovakav zapis je veoma intuitivan, i jasno pokazuje da mreža u procesu otklanjanja šuma iz slike x_t pokušava da proceni sliku $\hat{x}_\Theta(x_t)$ koja je što približnija početnoj slici x_0 . Naravno, zašumljena slika x_t nije dobijena na slučajan način, već tačno kontrolisanim zašumljivanjem početne slike u forward procesu na sledeći način:

$$x_t \sim q(x_t|x_0) = N(X_t|\sqrt{\bar{\alpha}_t}x_0, (1 - \bar{\alpha}_t)I) \quad (141)$$

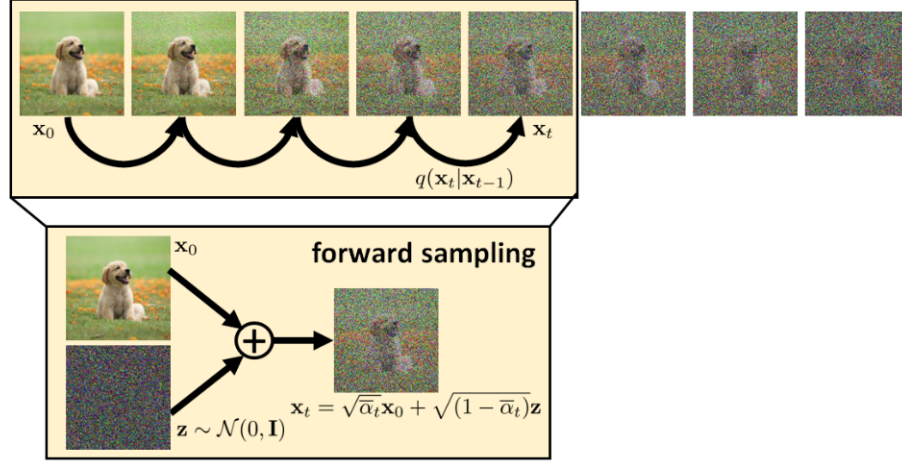
To jest:

$$x_t = \sqrt{\bar{\alpha}_t}x_0 + \sqrt{1 - \bar{\alpha}_t}z \quad (142)$$

gde je:

$$z \sim N(0, I) \quad (143)$$

Još jednom, forward proces bi vizuelno izgledao na sledeći način:



Slika 11: Forward proces difuzionog modela (preuzeto iz [2])

Funkcija gubitka nije ista za svaki difuzioni korak, to jest postoji parametar $\frac{1}{2\sigma_q^2(t)} \frac{(1-\alpha_t)^2 \bar{\alpha}_{t-1}}{(1-\bar{\alpha}_t)^2}$ koji kontroliše funkciju gubitka u svakom koraku otklanjanja šuma. Međutim, radi pojednostavljenja, ovu konstantu je moguće zanemariti, pošto ne utiče mnogo na krajnji rezultat.

Takođe suma $\sum_{t=1}^T$ se može apoksimirati izborom trenutka t iz uniformne raspodele $t \sim U(1, T)$.

Konačno, algoritam za treniranje difuzionog probabilističkog modela za procenu slike je sledeći:

1. Bira se vremenski trenutak iz raspodele $t \sim U(1, T)$.
2. Računa se zašumljena slika x_t po raspodeli $N(x_t | \sqrt{\alpha_t}x_0, (1 - \bar{\alpha}_t)I)$ to jest:

$$x_t = \sqrt{\alpha_t}x_0 + \sqrt{1 - \bar{\alpha}_t}z \quad (144)$$

gde je:

$$z \sim N(0, I) \quad (145)$$

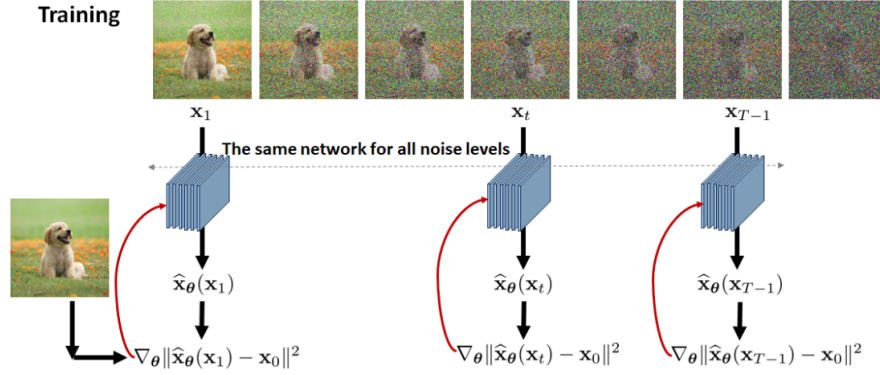
3. Model se trenira metodom gradijentnog spusta:

$$\nabla_{\Theta} \|\hat{x}_{\Theta}(x_t) - x_0\|^2 \quad (146)$$

4. Proces se ponavlja dok ne dođe do konvergencije funkcije gubitka.

Kao i bilo koja druga mreža, proces se može trenirati u batch-evima, što znači da se promena parametara (težina) mreže vrši tek nakon nekoliko odbiraka (slika), to jest nakon onoliko njih koliko ima u batch-u. Bitno je primetiti da se trenira jedna mreža za sve stepene zašumljivanja.

Proces treniranja modela je vizuelno predstavljen na sledećoj slici:



Slika 12: Treniranje difuzionog modela (preuzeto iz [2])

Kada je model istreniran, može se izvršiti zaključivanje uslovnih raspodela $p_{\Theta}(x_{t-1}|x_t)$ za svaki vremeski trenutak, odakle se mogu dobiti odbirci x_T, x_{T-1}, \dots, x_1 rekurzivno obrnutim procesom difuzije na sledeći način:

$$x_{t-1} \sim p_{\Theta}(x_{t-1}|x_t) = N(x_{t-1}|\mu_{\Theta}(x_t), \sigma_q^2(t)I) \quad (147)$$

$$= \mu_{\Theta}(x_t) + \sigma_q^2(t)z \quad (148)$$

$$= \frac{(1 - \bar{\alpha}_{t-1})\sqrt{\alpha_t}}{1 - \bar{\alpha}_t}x_t + \frac{(1 - \alpha_t)\sqrt{\bar{\alpha}_{t-1}}}{1 - \bar{\alpha}_t}\hat{x}_{\Theta}(x_t) + \sigma_q(t)z \quad (149)$$

gde je:

$$z \sim N(0, I) \quad (150)$$

Konačan algoritam zaključivanja u difuzionom probabilističkom modelu je sledeći:

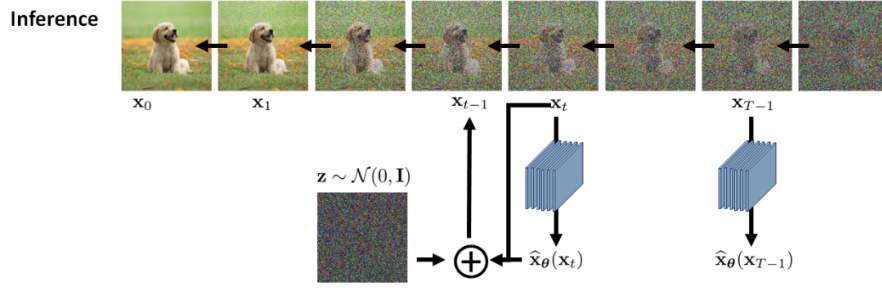
1. Kreće se od belog Gauss-ovog šuma $x_T \sim N(0, I)$.
2. Proces se ponavlja za svako $t = T, T-1, \dots, 1$.
3. Dobija se rezultat mreže $\hat{x}_{\Theta}(x_t)$.
4. Dobijaju se odbirci x_{t-1} uzorkovanjem na sledeći način:

$$x_{t-1} = \frac{(1 - \bar{\alpha}_{t-1})\sqrt{\alpha_t}}{1 - \bar{\alpha}_t}x_t + \frac{(1 - \alpha_t)\sqrt{\bar{\alpha}_{t-1}}}{1 - \bar{\alpha}_t}\hat{x}_{\Theta}(x_t) + \sigma_q(t)z \quad (151)$$

gde je:

$$z \sim N(0, I) \quad (152)$$

Proces zaključivanja je vizuelno prikazan na sledećoj slici:



Slika 13: Zaključivanje u difuzionom modelu (preuzeto iz [2])

5.6 Predviđanje šuma

Difuzionni modeli se takođe često treniraju na takav način da se ne predviđaju same slike već šum dodat na njih. Da bi se izveo matematički model toga potrebno je vratiti se na jednakost (41)

$$x_t = \sqrt{\alpha_t}x_0 + \sqrt{1 - \alpha_t}\epsilon_0 \quad (153)$$

i prepraviti je na sledeći način:

$$x_0 = \frac{x_t - \sqrt{1 - \alpha_t}\epsilon_0}{\sqrt{\alpha_t}} \quad (154)$$

Zamenom ovoga u jednakost (111) se dobija:

$$\mu_q(x_t, x_0) = \frac{(1 - \bar{\alpha}_{t-1})\sqrt{\alpha_t}x_t + (1 - \alpha_t)\sqrt{\alpha_{t-1}}x_0}{1 - \bar{\alpha}_t} \quad (155)$$

$$= \frac{(1 - \bar{\alpha}_{t-1})\sqrt{\alpha_t}x_t + (1 - \alpha_t)\sqrt{\alpha_{t-1}}\frac{x_t - \sqrt{1 - \bar{\alpha}_t}\epsilon_0}{\sqrt{\alpha_t}}}{1 - \bar{\alpha}_t} \quad (156)$$

što nakon sređivanja postaje:

$$\mu_q(x_t, x_0) = \frac{1}{\sqrt{\alpha_t}}x_t - \frac{1 - \alpha_t}{\sqrt{1 - \bar{\alpha}_t}\sqrt{\alpha_t}}\epsilon_0 \quad (157)$$

Po ugledu na ovo matematičko očekivanje, $\mu_\Theta(x_t)$ postaje:

$$\mu_\Theta(x_t) = \frac{1}{\sqrt{\alpha_t}}x_t - \frac{1 - \alpha_t}{\sqrt{1 - \bar{\alpha}_t}\sqrt{\alpha_t}}\hat{\epsilon}_0(x_t) \quad (158)$$

Zamenom ovoga u jednakost (127) se za donju granicu dokaza dobija:

$$\text{ELBO}_\Theta(x) = - \sum_{t=1}^T E_{q(x_t|x_0)} \left[\frac{1}{2\sigma_q^2(t)} \frac{(1 - \alpha_t)^2 \bar{\alpha}_{t-1}}{(1 - \bar{\alpha}_t)^2} \|\hat{\epsilon}_\Theta(x_t) - \epsilon_0\|^2 \right] \quad (159)$$

Sada se algoritam treniranja ne mreže zasniva na proceni slike već šuma dodatog na tu sliku i izgleda na sledeći način:

1. Bira se vremenski trenutak iz raspodele $t \sim U(1, T)$.
2. Računa se zašumljena slika x_t po raspodeli $N(X_t | \sqrt{\alpha_t}x_0, (1 - \bar{\alpha}_t)I)$ to jest:

$$x_t = \sqrt{\alpha_t}x_0 + \sqrt{1 - \bar{\alpha}_t}z \quad (160)$$

gde je:

$$z \sim N(0, I) \quad (161)$$

3. Model se trenira metodom gradijentnog spusta:

$$\nabla_{\Theta} ||\hat{\epsilon}_{\Theta}(x_t) - \epsilon_0||^2 \quad (162)$$

4. Proces se ponavlja dok ne dođe do konvergencije funkcije gubitka.

Kada je model istreniran, može se izvršiti zaključivanje uslovnih raspodela $p_{\Theta}(x_{t-1}|x_t)$ za svaki vremeski trenutak, odakle se mogu dobiti odbirci x_T, x_{T-1}, \dots, x_1 rekursivno obrnutim procesom difuzije na sledeći način:

$$x_{t-1} \sim p_{\Theta}(x_{t-1}|x_t) = N(x_{t-1}|\mu_{\Theta}(x_t), \sigma_q^2(t)I) \quad (163)$$

$$= \mu_{\Theta}(x_t) + \sigma_q^2(t)z \quad (164)$$

$$= \frac{1}{\sqrt{\alpha_t}}x_t - \frac{1 - \alpha_t}{\sqrt{1 - \bar{\alpha}_t}\sqrt{\alpha_t}}\hat{\epsilon}_{\Theta}(x_t) + \sigma_q(t)z \quad (165)$$

$$= \frac{1}{\sqrt{\alpha_t}}(x_t - \frac{1 - \alpha_t}{\sqrt{1 - \bar{\alpha}_t}}\hat{\epsilon}_{\Theta}(x_t)) + \sigma_q(t)z \quad (166)$$

gde je:

$$z \sim N(0, I) \quad (167)$$

Konačan algoritam zaključivanja u difuzionom probabilističkom modelu pri zaključivanju šuma je sledeći:

1. Kreće se od belog Gauss-ovog šuma $x_T \sim N(0, I)$.
2. Proces se ponavlja za svako $t = T, T - 1, \dots, 1$.
3. Dobija se rezultat mreže $\hat{\epsilon}_{\Theta}(x_t)$.
4. Dobijaju se odbirci x_{t-1} uzorkovanjem na sledeći način:

$$x_{t-1} = \frac{1}{\sqrt{\alpha_t}}(x_t - \frac{1 - \alpha_t}{\sqrt{1 - \bar{\alpha}_t}}\hat{\epsilon}_{\Theta}(x_t)) + \sigma_q(t)z \quad (168)$$

gde je:

$$z \sim N(0, I) \quad (169)$$

Ovakav algoritam je predložen u radu [3]:

Algorithm 1 Training	Algorithm 2 Sampling
1: repeat 2: $\mathbf{x}_0 \sim q(\mathbf{x}_0)$ 3: $t \sim \text{Uniform}(\{1, \dots, T\})$ 4: $\epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ 5: Take gradient descent step on $\nabla_{\theta} \epsilon - \epsilon_{\theta}(\sqrt{\bar{\alpha}_t}\mathbf{x}_0 + \sqrt{1 - \bar{\alpha}_t}\epsilon, t) ^2$ 6: until converged	1: $\mathbf{x}_T \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ 2: for $t = T, \dots, 1$ do 3: $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ if $t > 1$, else $\mathbf{z} = \mathbf{0}$ 4: $\mathbf{x}_{t-1} = \frac{1}{\sqrt{\alpha_t}} \left(\mathbf{x}_t - \frac{1 - \alpha_t}{\sqrt{1 - \bar{\alpha}_t}} \epsilon_{\theta}(\mathbf{x}_t, t) \right) + \sigma_t \mathbf{z}$ 5: end for 6: return \mathbf{x}_0

Slika 14: Algoritmi treniranja i zaključivanja (preuzeto iz [3])

6 Arhitektura mreže

Tip mreže koji je treniran ranije objašnjenim algoritmom je U-Net [1].

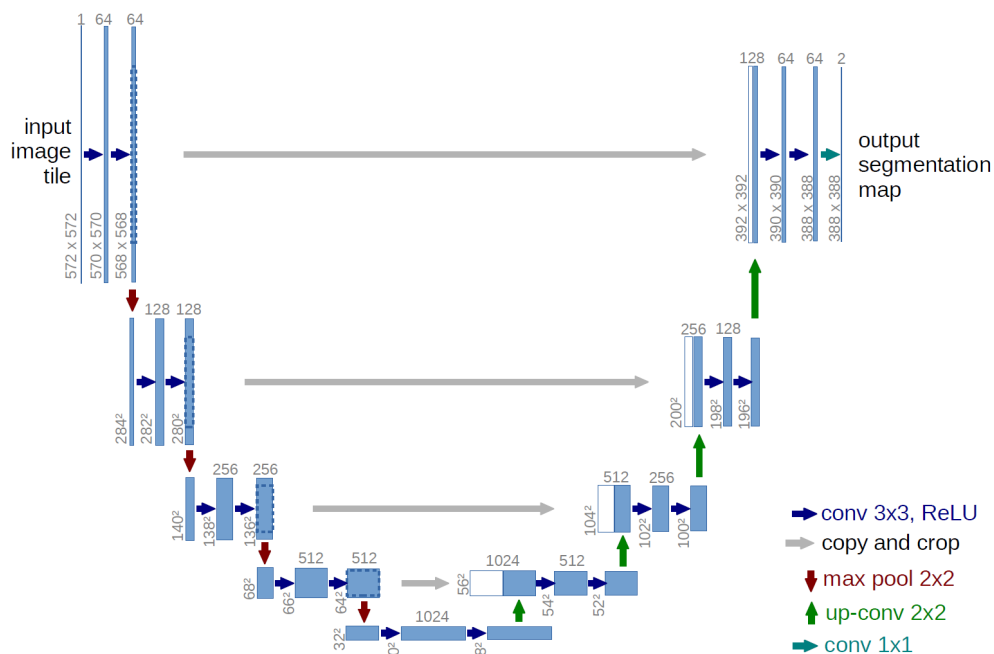
Naime, U-Net je tip autoenkoderske mreže koja se sastoji iz dela koji se naziva enkoder i dela koji se naziva dekodek.

Cilj enkodera je da pomoću nekoliko enkoderskih blokova postepeno polovi rezoluciju slike, kako bi se izdvojile najbitnije karakteristike i obeležja. Enkoderski blok se sastoji iz nekoliko delova u koje spadaju konvolucionni slojevi, normalizacija batch-a i MaxPool sloj. Konvolucionni sloj je skup 3x3 kernela koji ulaze u konvoluciju sa samim ulaznim podacima, i elementi tih kernela su upravo ono što se uči u okviru treniranja mreže. Normalizacija batch-a se odnosi na to da se svi dobijeni podaci normalizuju, kako neki od njih ne bi postali suviše dominantni u odnosu na ostale. MaxPool sloj se odnosi na kernel veličine 2x2 koji iz svake grupe od četiri piksela sa slika izdvaja onaj piksel sa najvećom vrednošću. Na taj način se efektivno izdvajaju najbitnija obeležja i polovi se rezolucija svih slika na kraju svakog enkoderskog bloka.

Cilj dekodekera je da obrne postupak enkodera i da sliku vrati na početnu rezoluciju, čuvajući izdvojena najbitnija obeležja, koja su produkt enkodera. Dekodek se takođe sastoji iz dekoderskih blokova koji imaju sličnu strukturu enkoderskim blokovima. Jedina razlika u odnosu na enkoderske blokove je to što se umesto MaxPool sloja koristi specijalan konvolucionni sloj koji povećava rezoluciju slike (u konkretnom slučaju duplira). Pored toga, ulaz u dekoderske blokove ne predstavlja samo izlaz iz prethodnog bloka, već se na njega nadovezuje i izlaz enkoderskog bloka istog nivoa, i ovaj tip nadovezivanja se u literaturi naziva skip konekcija. Cilj skip konekcije je da spreči takozvanu pojavu vanishing gradients koja se odnosi na saturaciju parametara mreže. Takođe je želja da se u procesu dekodovanja ne izgubi previše informacija i zato se nadovezuju i izlazi enkodera.

Takođe, pri smanjivanju rezolucije, u svakom narednom enkoderskom bloku ima dva puta više konvolucionih kernela nego u prethodnom, dok pri povećanju rezolucije u svakom narednom bloku ima dva puta manje kernela nego u prethodnom. Pored toga, u svakom konvolucionom sloju je korišćena ReLU aktivaciona funkcija.

Struktura U-Net-a se može videti na sledećoj slici:



Slika 15: Struktura U-Net-a (preuzeto iz [1])

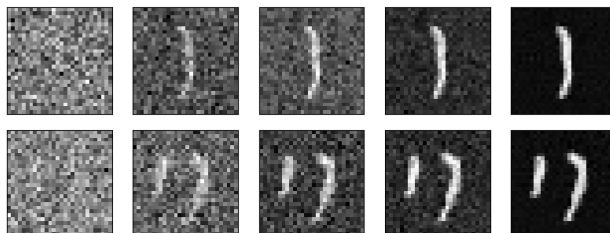
7 Rezultati i diskusija

U cilju treniranja modela, ulazne slike iz MNIST baze podataka su zašumljene do maksimalnog vremenskog trenutka $T=300$. Korišćena veličina jednog batch-a je 512, i treniranje je rađeno prvo u 50, a zatim u 100 epoha. Takođe je korišćen korak učenja 0.001, kao i *Adam* optimizacija, gde je implementacija ovako definisanog modela data na narednoj putanji:

<https://colab.research.google.com/drive/1ghj0OV9fMd8dD6vz5c867P0PBwibm0BX?usp=sharing>.

Nakon treniranja modela je vršen rekurzivni proces zaključivanja, kojim je iz zašumljene slike procesom obrnute difuzije dobijena procena ulazne slike, kao i procena zašumljene slike u svim vremenskim trenucima.

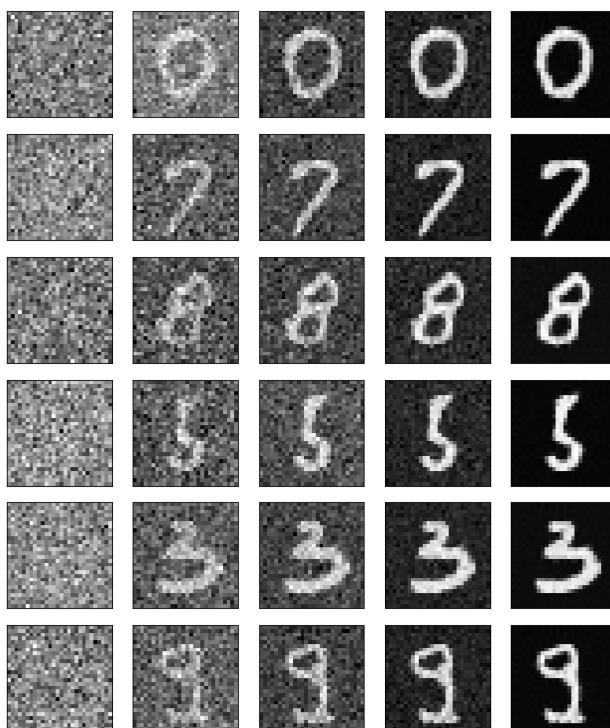
Proces obrnute difuzije, nakon 50 epoha treniranja mreže je dat na sledećim slikama:



Slika 16: Rezultati nakon 50 epoha

Na prethodnim slikama je predstavljen proces obrnute difuzije u vremenskim trenucima $t = 300, 100, 50, 20$ kao i procenjena ulazna slika.

Kao što se može primetiti, konačni rezultati nisu idealni, ali se na datim slikama razaznaju oblici. Na drugoj priloženoj slici se pak ne razaznaje nijedna cifra, čemu razlog može da bude ili loše napisana cifra, ili, naravno, nedovoljno treniran model. Iz tog razloga, u cilju boljih rezultata je treniranje modela nastavljeno dodatnih 50 epoha, i dobijeni rezultati backward procesa su sledeći:





Slika 17: Rezultati nakon 100 epoha

Kao što se na prethodnim slikama može primetiti, rezultati su prilično zadovoljavajući. Funkcija gubitka opada do vrednosti 0.1279, i na primerima iz test skupa se nakon procesa obrnute difuzije jasno razaznaju rukom napisane cifre.

Naravno, treba uzeti u obzir da su podaci nad kojima je vršeno treniranje veoma jednostavni i male rezolucije. Iz tog razloga je trenirani model ograničenih dimenzija, što bi svakako bilo neophodno da se unapredi ukoliko bi se radilo sa složenijim podacima.

7.1 Unapređenja

Pored do sada opisane arhitekture i elemenata difuzionih modela postoji dosta unapređenja, kao i modifikacija komponenti kako forward, tako i backward procesa.

Jedan od elemenata u kome ima prostora za modifikacije je šum u forward procesu. Preciznije, moguće je varirati tip šuma koji je iterativno dodavan na ulazne podatke u forward procesu, kao i intenzitet šuma koji je dodat u svakom od koraka.

Količina šuma koja je dodata u svakom koraku je određena parametrom β_t , kao što je ranije objašnjeno. Način na koj se taj parametar menja postiže ravnotežu između istraživanja i eksploatacije. Istraživanje opisuje sposobnost generalizacije na podacima koji nisu viđeni tokom treninga, dok eksploatacija predstavlja sposobnost modela da konvergira u treniranju nad poznatim podacima. Što je veća količina šuma koja se koristi, brže se uništavaju strukture podataka, i obrnuto. S jedne strane, dovoljna količina šuma je neophodna da bi se podstaklo istraživanje kako bi bila dobra generalizacija na nepoznatim podacima, dok prekomerni šum može dovesti do suboptimalne konvergencije ili nekonvergencije modela. S druge strane, premalo šuma podstiče eksploataciju kako bi model konvergirao, ali to narušava generalizaciju.

Prilikom dizajna količine šume u svakom koraku se često uzimaju u obzir i sami podaci, na primer njihova dimenzionalnost ili maksimalno euklidsko rastojanje odabrak trening skupa, kao i redundantnost i složenost podataka. Na primer, veće slike mogu zahtevati više šuma [5].

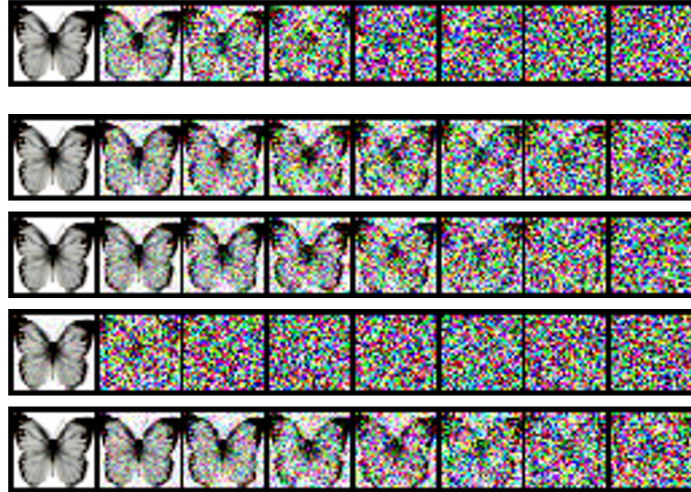
Određene implementacije modela, pored gore objašnjenih parametara, uče i raspored dovanja šuma, β_t , i u tom slučaju šum može biti drugačiji i forward i backward procesu. Sa druge strane, većina implementacija zadržava analitički definisane odblike parametra β_t , gde je pokazano da linearne ili eksponencijalne promene parametra β_t kroz vreme deformišu ulazne podatke nešto brže nego što je potrebno. Kao posledica toga koriste se druge funkcije, poput kosinusne funkcije, sigmoid funkcije ili jednostavne linearne funkcije. U radu [6] je predložena promena parametra α_t po jednostavnoj linearnoj funkciji koja je definisana na sledeći način:

$$\alpha_t = 1 - t \quad (170)$$

gde je $t \sim U(0, 1)$, i ovakava definicija odgovara kontinualnoj formulaciji difuzionih modela. Pandam ovakvoj definiciji u diskretnoj formulaciji difuzionih modela, koja je do sada obrađivana u radu bi izgledao na sledeći način:

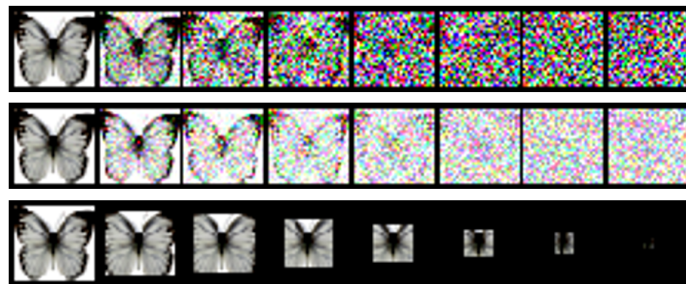
$$\alpha_t = 1 - \frac{t}{T} \quad (171)$$

Sa druge strane linearna promena parametra β_t u vremenu je definisana kao $\beta_t = at + b$. Na narednoj slici su vizuelno prikazani efekti navedenih varijanti dodavanja šuma:



Slika 18: Poređenje promene dodavanja šuma na različite načine - 1. linearna funkcija, 2. jednostavna linearna funkcija, 3. kosinusna funkcija, 4. eksponencijalna funkcija, 5. sigmoid funkcija (preuzeto iz [5])

Pored načina dodavanja šuma, može se varirati i sam tip šuma koji se dodaje. Kao i u ovom radu, najčešće korišćeni je Gauss-ov šum sa jediničnom kovarijacionom matricom (izotropski Gauss-ov šum), dok pojedini radovi koriste modifikaciju Gauss-ovog šuma sa kovarijacionom matricom koja nije jedinična (anizotropski Gauss-ov šum) kao i kombinaciju oba. Kao alternativa Gauss-ovom šumu se koristi šum urorkovan iz gama raspodele, koja ima više stepeni slobode, kao i različiti tipovi blagih korupcija koje maskiraju slike posebno dizajniranim kernelima. Na narednoj slici su prikazani različiti tipovi korišćenih šumova i način na koji oni deformišu slike:



Slika 19: Poređenje različitih tipova šuma-1. Gauss-ov šum, 2. šum sa gama raspodelom, 3. blage korupcije (preuzeto iz [5])

Pored ovoga vredi napomenuti da arhitektura modela tipa U-Net nije jedina arhitektura korišćena u difuzionim modelima, ali jeste do sada nakorišćenija i najprilagodljivija arhitektura za ovakav tip problema. Pored U-Net arhitekture se često koriste i transformeri koji su, poput U-Net-a, mreža koja se sastoji od enkoderskog i dekoderskog dela, međutim, za potrebe difuzionih modela se u tom slučaju u transformere dodaju neki elementi U-Net-a, poput skip konekcija. Takođe, neke implementacije, pored matematičkog očekivanja uslovnih raspodela obrnutog procesa difuzije, uče i njihovu kovarijacionu matricu, ili je modeluju različitim analitičkim oblicima [5].

8 Zaključak

U prethodnom radu je detaljno predstavljena teorijska osnova difuzionih probablističkih modela za uklanjanje šuma. Objasnen je forward proces, u kome se ulazni podaci iterativno zašumljuju, dok ne postanu što približniji belom Gauss-ovom šumu, kao i backward proces obrnute difuzije, u kome se varijacionim zaključivanjem dolazi do uslovnih raspodela za podatke iz prethodnog vremenskog trenutka. Takođe je objasnen postupak formiranja i treniranja modela čija je uloga procena gore navedenih aposteriornih raspodela.

Prikazani su i dobijeni rezultati, koji su za potrebe rada zadovoljavajući, međutim postoji dosta prostora za unapređenje ukoliko bi želja bila rad sa složenijim podacima. Pored toga su objašnjene i moguće modifikacije i unapređenja korišćena u različitim implementacijama.

Difuzioni modeli su se pokazali kao dobar alat rešavanja različitih tipova problme poput popunjavanja praznina na slikama, generisanja slika iz teksta, segmentacije, generisanja govora, sinteze muzike, poboljšanja audio signala. Pored toga postaju popularni u oblastima sinteze molekula, sinteze ljudskog pokreta, predikcije trajektorije, simulacijama, i drugim oblastima i predikcije su da će postajati sve više korišćeni.

9 Literatura

- [1] Lilian Weng "What are Diffusion Models?" dostupno na lokaciji:<https://lilianweng.github.io/posts/2021-07-11-diffusion-models/>
- [2] Stanley H. Chan "Tutorial on Diffusion Models for Imaging and Vision" arXiv:2403.18103 [cs.LG]
- [3] Jonathan Ho, Ajay Jain, Pieter Abbeel Denoising Diffusion Probabilistic Models" arXiv:2006.11239 [cs.LG]
- [4] Machine Learning & Simulation. (2021, maj 11). Variational Inference | Evidence Lower Bound (ELBO) | Intuition & Visualization [Video zapis]. YouTube. <https://www.youtube.com/watch?v=HxQ94L8n0vU>
- [5] Ziyi Chang, George Alex Koulieris, Hubert P. H. Shum "On the Design Fundamentals of Diffusion Models: A Survey" arXiv:2306.04542 [cs.LG]
- [6] Ting Chen "On the Importance of Noise Scheduling for Diffusion Models" arXiv:2301.10972 [cs.CV]