

# Метод генерації синтетичних зображень на основі генеративно-дискримінаційної нейромережі

## **Виконав:**

студент 4 курсу  
групи КА-51  
Сайног О. М.

## **Керівник:**

к. ф.-м. н., доцент  
Каніовська І. Ю.

# Актуальність дослідження

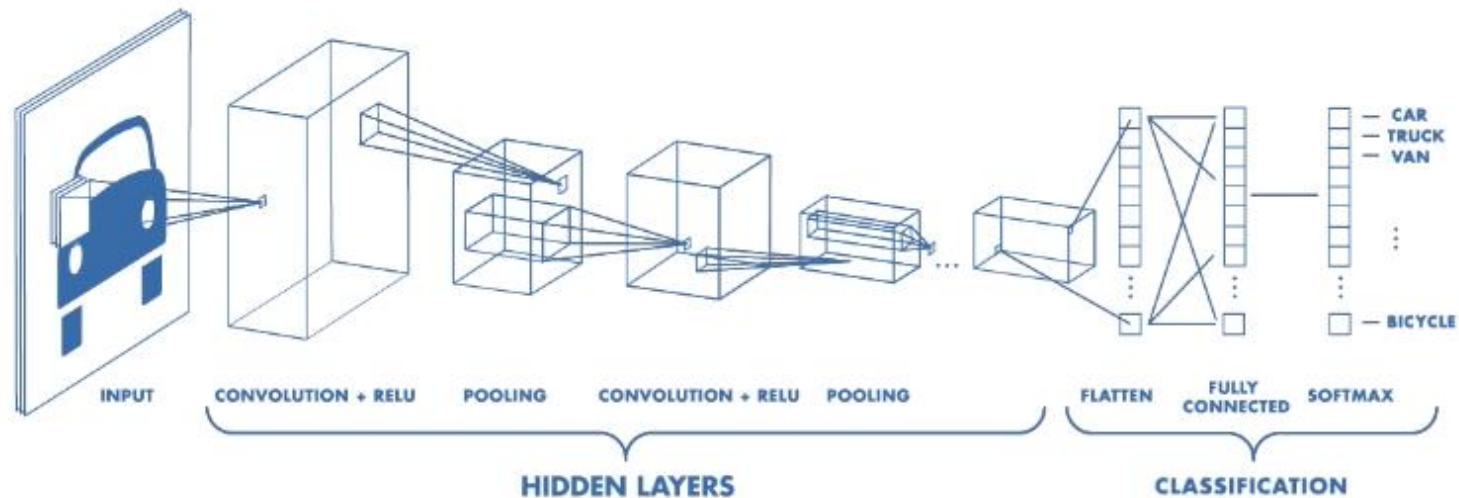
- Ідея змагання двох мереж, що навчаються, вважаються найперспективнішою ідеєю для розвитку методів глибинного навчання.
- Синтетичні зображення активно використовуються для покращення навчання вже існуючих моделей розпізнавання, класифікації чи ідентифікації об'єктів.
- Згенеровані зображення можуть використовуватися для створення елементів віртуальної реальності, що набирає темпи і обсяги розвитку, та впроваджується в повсякдення.
- Деякі з розглянутих моделей можна використовувати як стилізатори, що значно збільшує варіації для нових зображень.

- **Об'єкт дослідження**  
синтетичні зображення
- **Предмет дослідження**  
методи генерації зображень на основі варіаційних автоматичних кодувальників (ВАК) та генеративно-дискримінаційних нейронних мереж (ГДМ).
- **Мета дослідження**  
проаналізувати предмет дослідження, реалізувати деякі архітектури ГДМ, провести їх тестування та порівняти результати.

# Постановка задачі

1. Створення програмного продукту для генерації зображень відповідного типу.
2. В результаті роботи програми порівняти результати генерації на основі:
  - ❖ умовного варіаційного автоматичного кодувальника
  - ❖ умовної генеративно-дискримінаційної нейромережі
  - ❖ умовної генеративно-дискримінаційної нейромережі з вбудованим умовним варіаційним автоматичним кодувальником.
3. Додати пояснення різниці генерації різними методами та обґрунтувати, чим поєднання двох моделей є найкращим з точки зору візуалізованого результату.

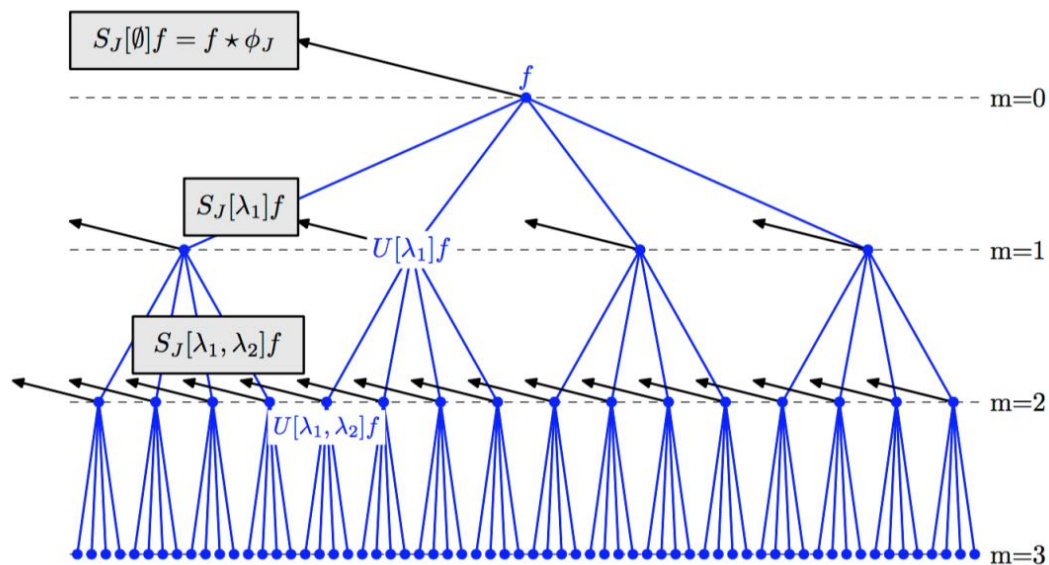
# Згорткові нейронні мережі



Вперше згорткові нейромережі були запропоновані Яном Ле Куном як біологічно натхнена мережа, працеспроможність якої було доведено емпіричним шляхом.

Основою стала операція згортки вхідного сигналу з певним тензором фільтрів, коефіцієнти якого оптимізуються в процесі навчання.

# Згорткові нейронні мережі

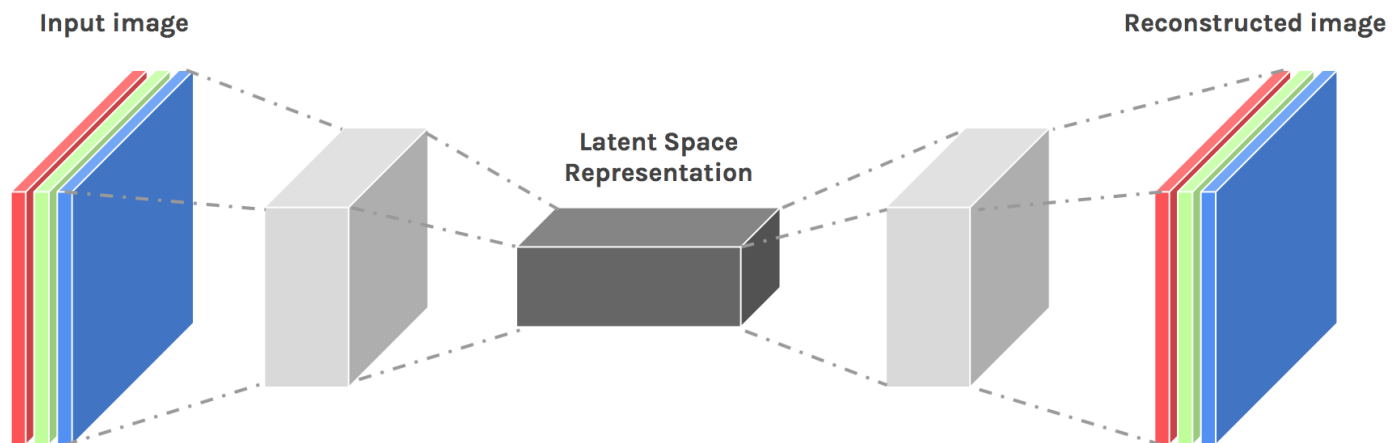


Стефан Маллат надав математичне обґрунтування введеної раніше архітектури. Спершу було розглянуто неперервне вейвлет-перетворення розсіювання для одного каналу згортки, що зображено на малюнку.

Наступним кроком стала заміна вимог скорочень та інваріантів до трансляцій скорочень вздовж адаптивних груп локальних симетрій.

Потім, вейвлети замінюються адаптивними фільтрами ваги, подібними до моделей глибокого навчання.

# Варіаційний автокодувальник (ВАК)

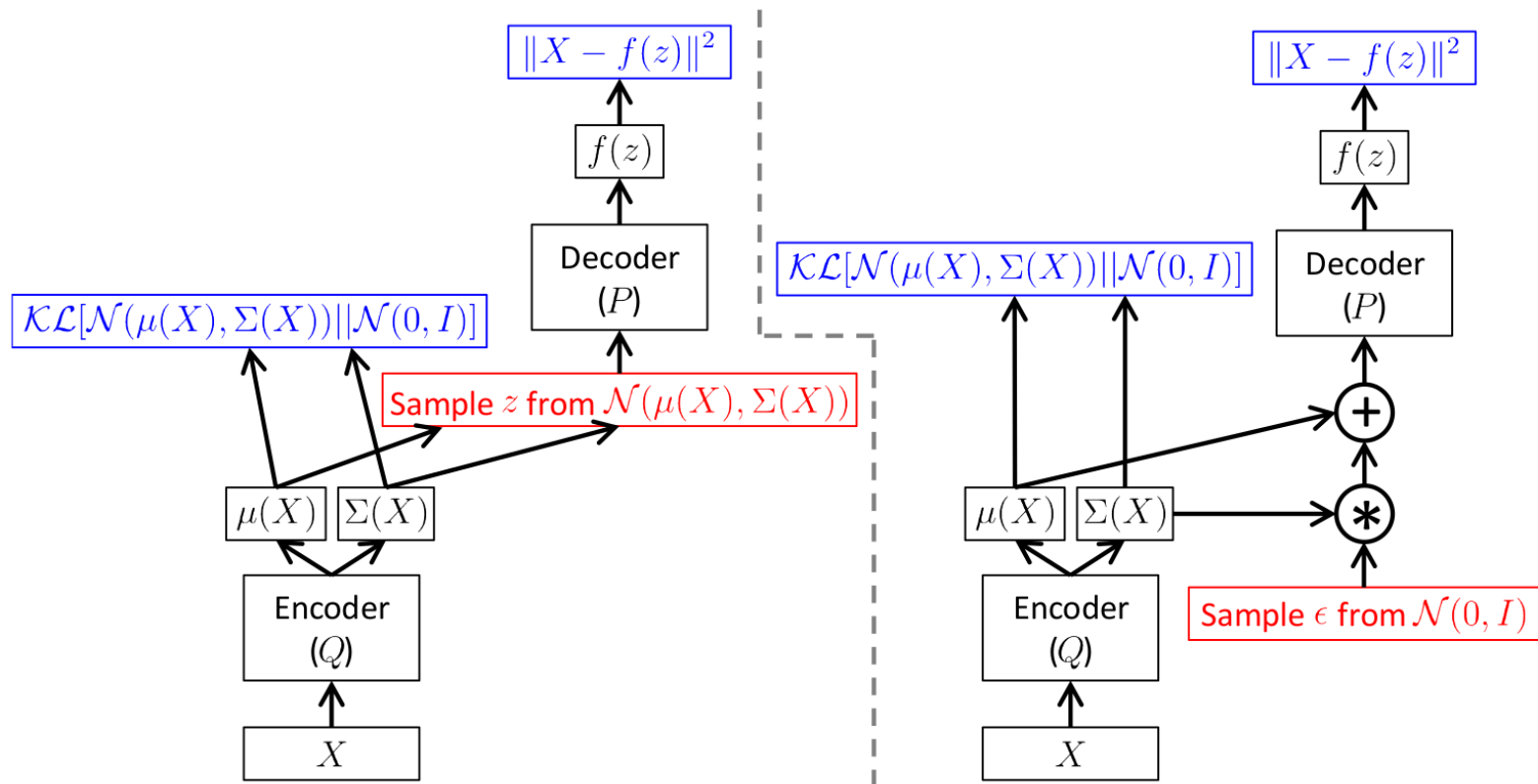


Ідея автоматичного кодування полягає в побудові двох мереж:

1. Кодувальник – перетворює вхідний сигнал в елемент прихованого простору (код).
2. Декодувальник - відновлює вхідний сигнал з його коду.

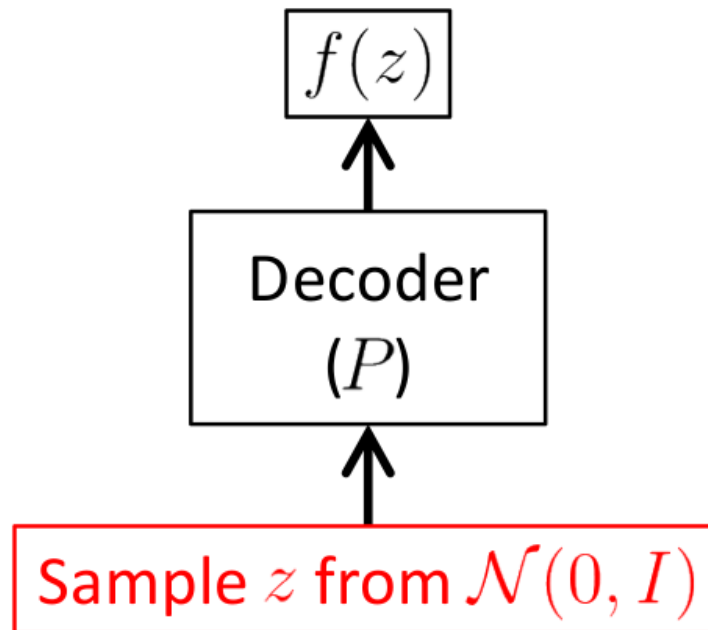
Мета варіаційного автокодувальника - створення наперед заданого розподілу в просторі прихованих змінних (кодів). Зазвичай використовують багатовимірний нормальний розподіл.

# Схема навчання ВАК





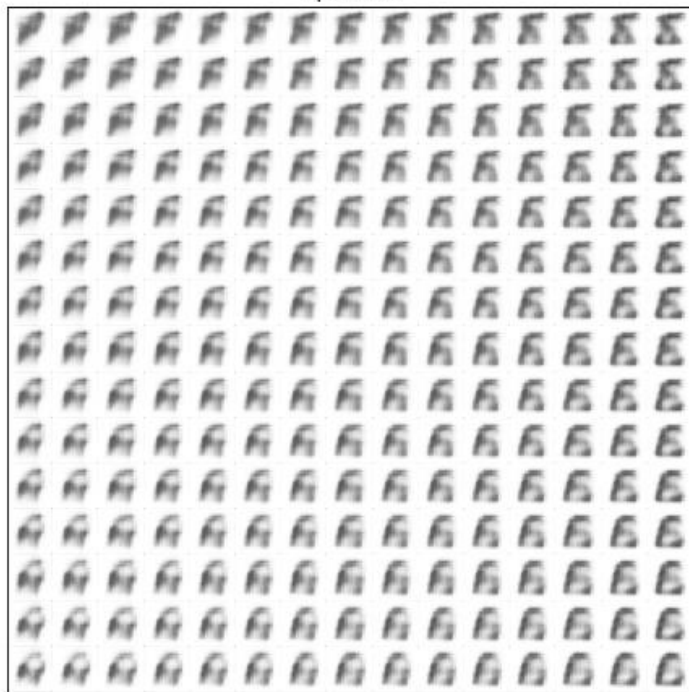
# Генераційна модель



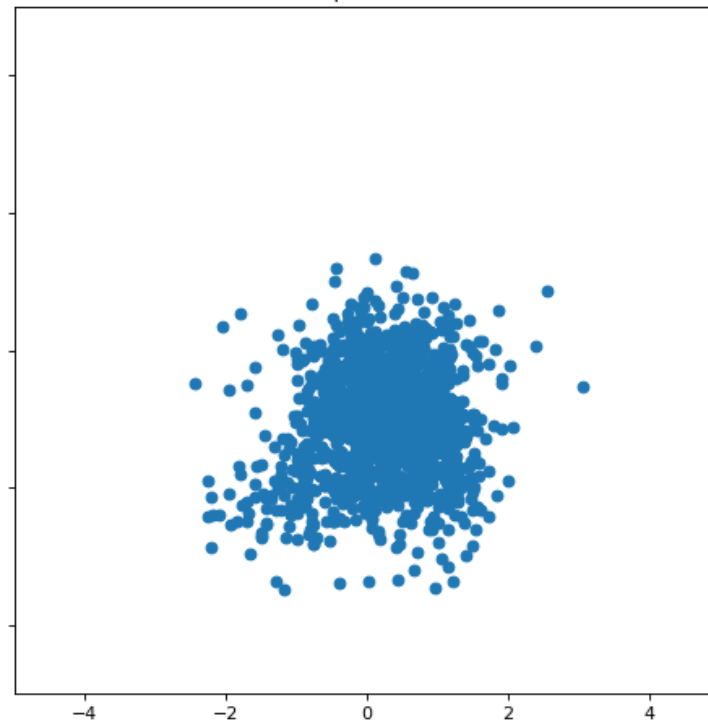
В результаті маємо повноцінну генеративну модель в вигляді декодувальника.

# Процес навчання ВАК

Label: 5  
Epoch: 0



Label: 5  
Epoch: 0



# Результати генерації ВАК

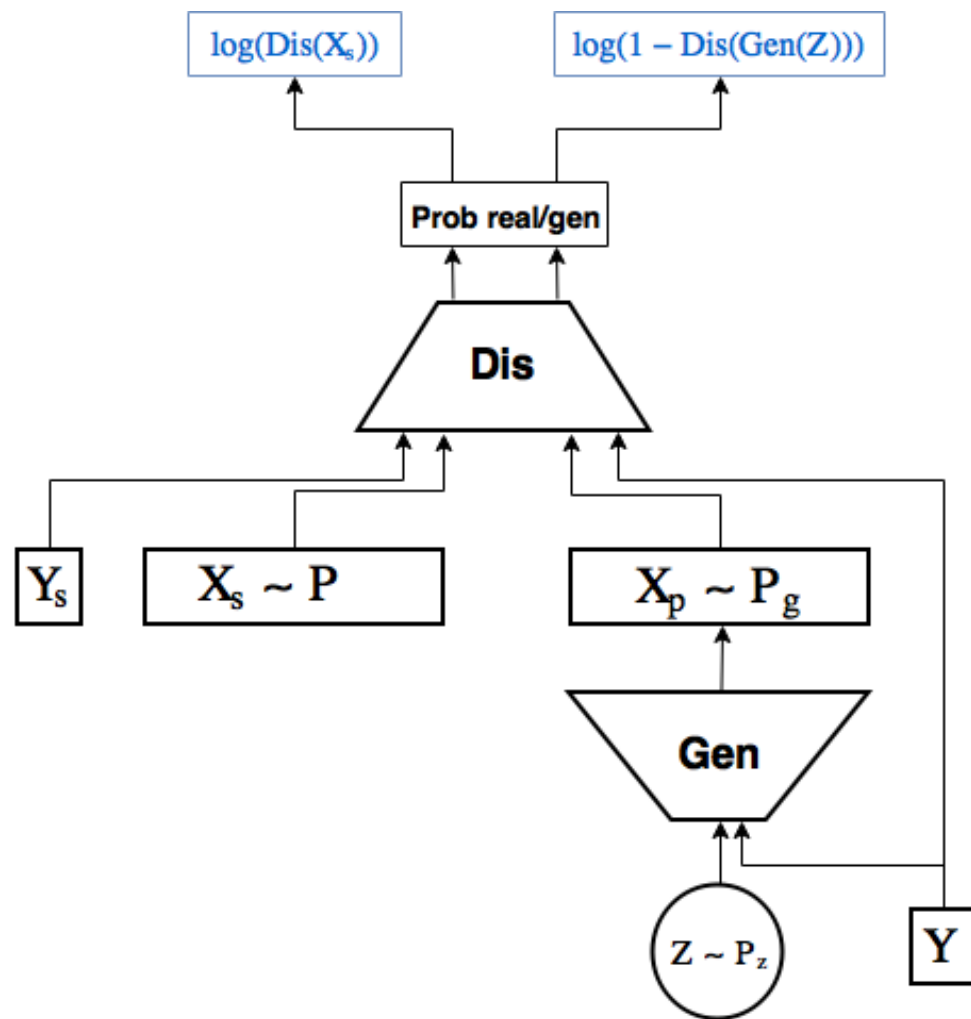
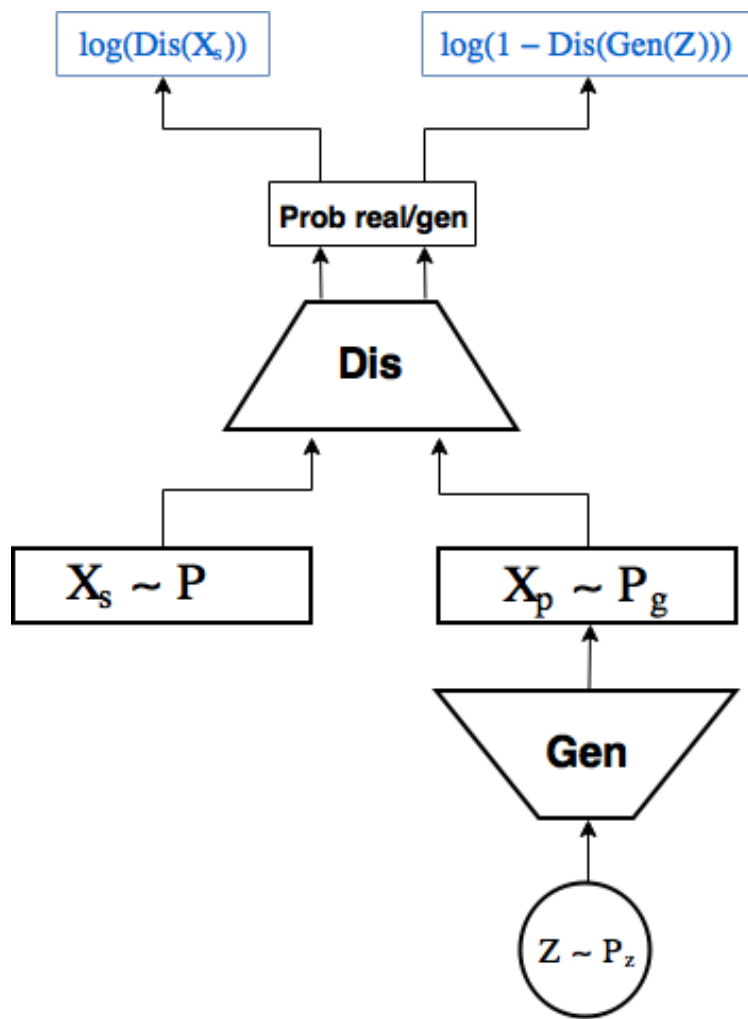


# Стилiзацiя вхiдних зображень за допомогою ВАК



# Генеративно-дискримінаційна нейронна мережа (ГДМ)

ГДМ є найкращими на даний час генеративними ідеями за рахунок протиставлення двох нейромереж – генератора та дискримінатора. Основною задачею дискримінатора – визначення чи є зображення реальним чи згенерованим, чим найкраще наближує синтетичні зображення до природних.



# Процес навчання ГДМ

Генератор та дискримінатор навчаються по черзі як одна мережа ( $k$  циклів навчання дискримінатора з параметрами  $\theta_d$  та один цикл навчання генератора з параметрами  $\theta_g$ ).

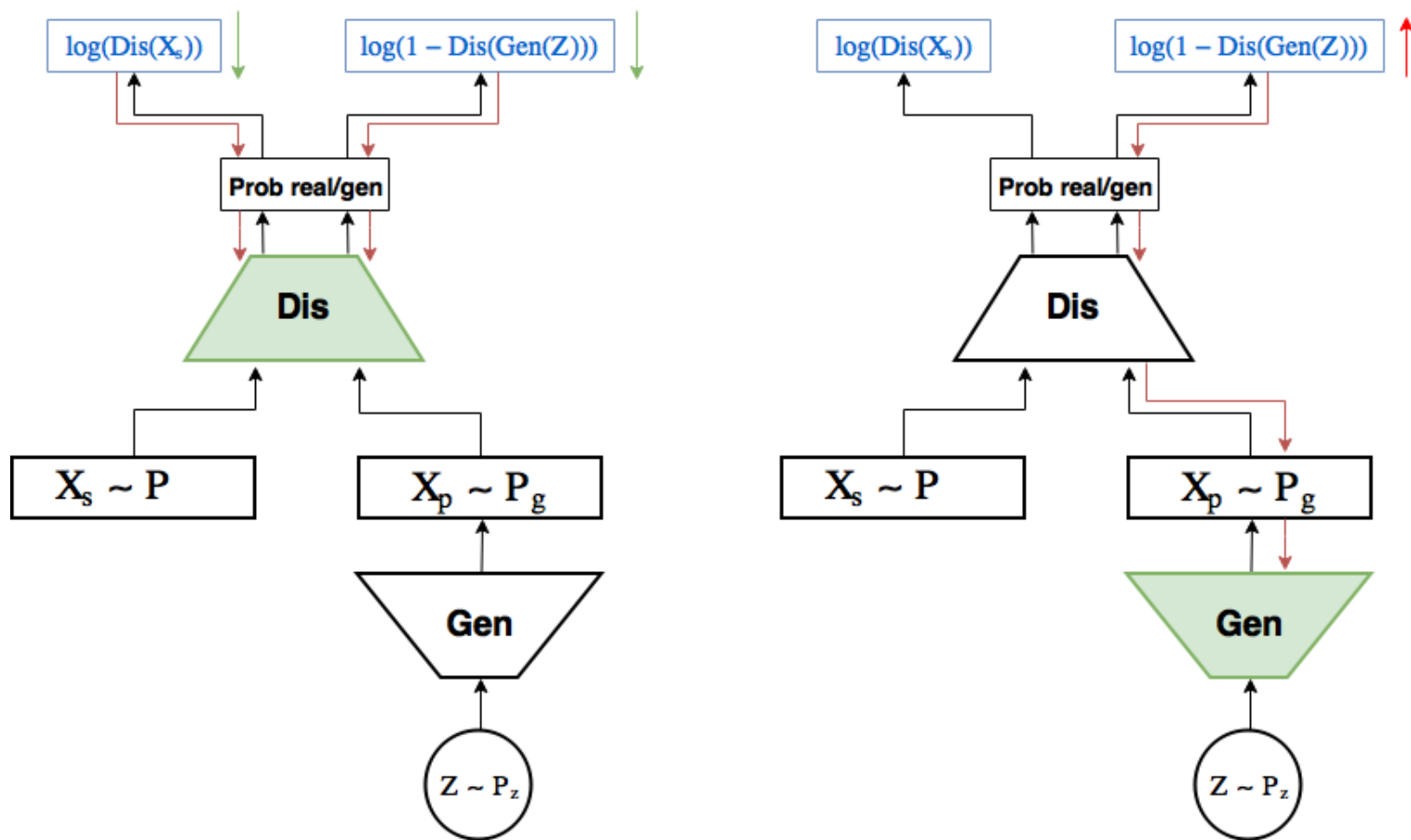
Дискримінатор мінімізує бінарну перехресну ентропію:

$$\theta_d := \theta_d - \nabla_{\theta_d} \left( \ln(D(X_s)) + \ln(1 - D(G(Z))) \right)$$

Генератор максимізує логарифм ймовірності того, що дискримінатор визначить згенероване зображення як реальне:

$$\theta_g := \theta_g + \nabla_{\theta_g} \ln(1 - D(G(Z)))$$

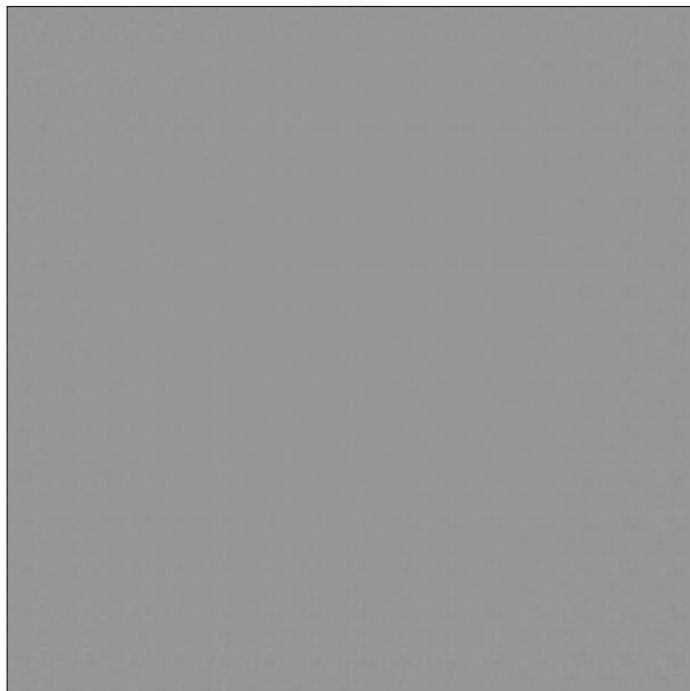
# Процес навчання ГДМ



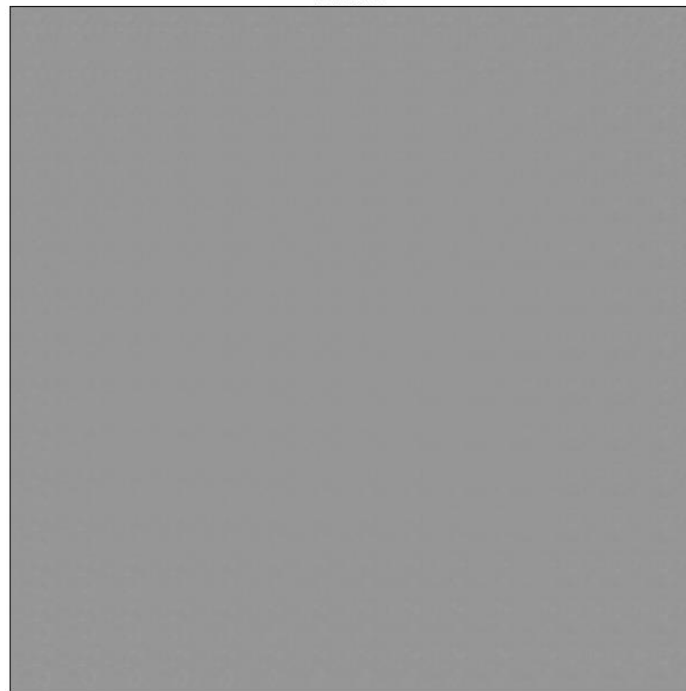


# Процес навчання ГДМ

Label: 11  
Batch: 0



Label: 19  
Batch: 0

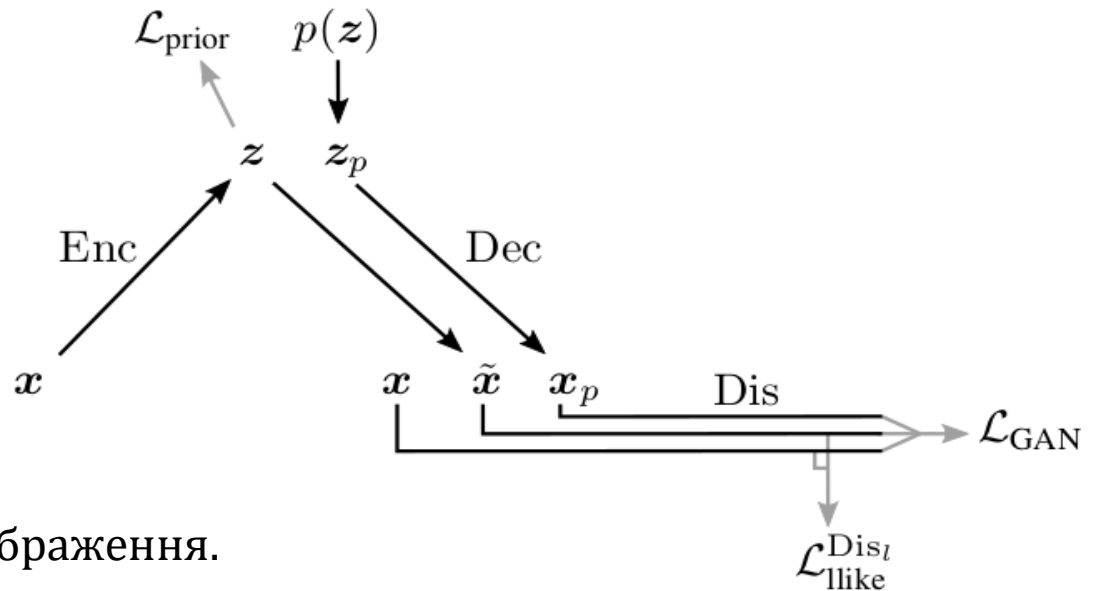


# Поєднання ВАК та ГДМ

- Використовуючи умовний варіаційний автокодувальник можна генерувати зображення заданого класу, а також зображення одного класу зі стилізацією іншого. За допомогою умовної ГДМ генерувати стилізовані зображення неможливо. Поєднання обох архітектур в одну зводиться до додання в архітектуру автокодувальника дискримінатора.

Позначення:

- $x$  – об'єкт з  $P(x)$ .
- $z_p$  – обраний код з  $P(z)$ .
- $x_p$  – згенероване з  $z_p$  зображення.
- $\tilde{x}$  – об'єкт відновлений з  $x$ .
- $\mathcal{L}_{prior} = KL[Q(z|x)||P(z)]$  – функція помилки, що наближає значення кодувальника на  $P(x)$  до  $P(z)$ .
- $\mathcal{L}_{llike}^{Disl} = L_d(d_l(x), d_l(\tilde{x}))$  – функція помилки між активаціями  $l$ -го шару дискримінатора  $D$  на реальному  $x$  та відновленому  $\tilde{x} = f_d(Q(x))$ .
- $\mathcal{L}_{GAN} = \ln(D(x)) + \ln(1 - D(f_d(z))) + \ln(1 - D(f_d(Q(x))))$  – перехресна ентропія між дійсним розподілом класів реальних чи згенерованих об'єктів і розподілом ймовірностей прогнозованих дискримінатором.



# Навчання загальної моделі

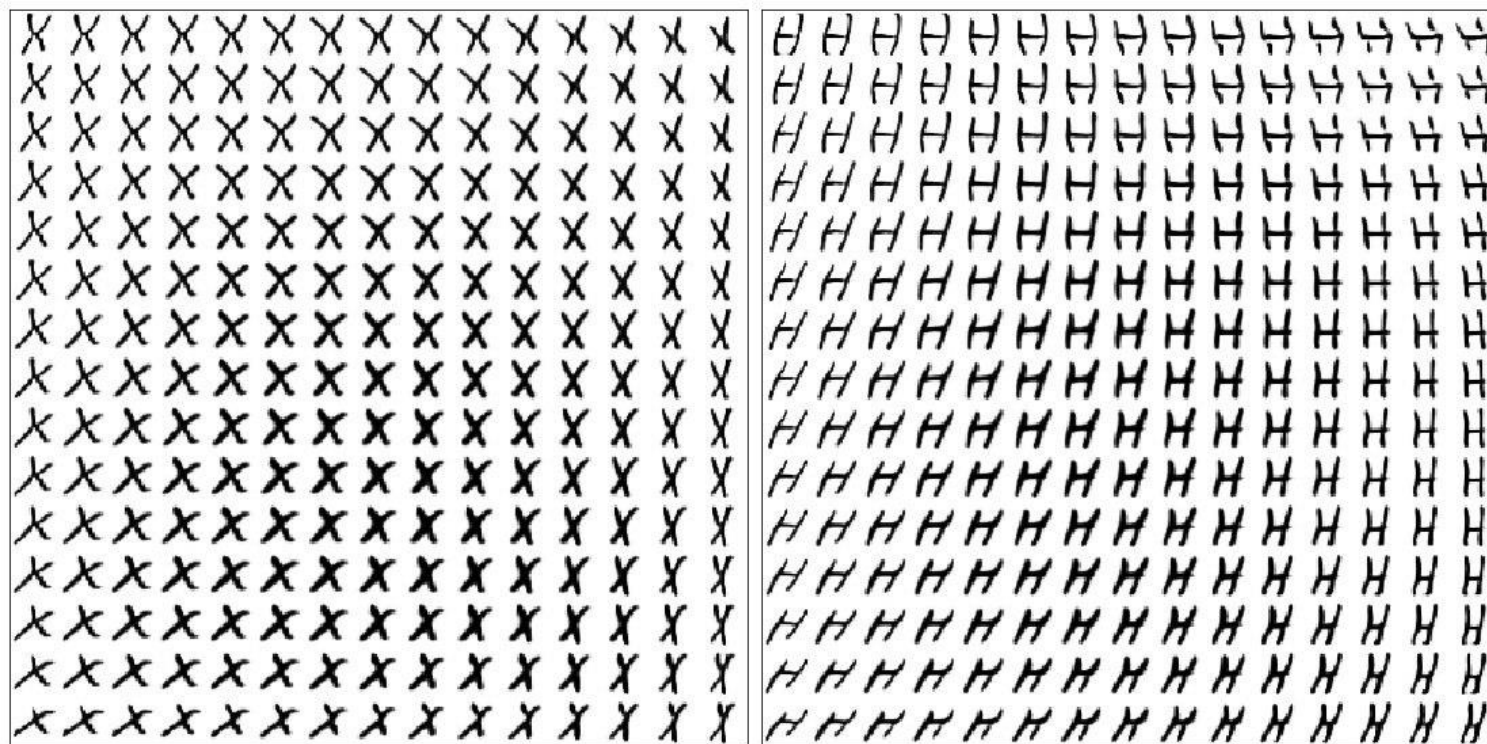
Очевидно, що навчання не може відбуватися одночасно, а кожна з під мереж повинна мінімізувати лише деякі функції помилки (дискримінатор не може мінімізувати  $\mathcal{L}_{llike}^{Disl}$ , бо це призведе до однакової активації усіх шарів).

Використаємо наступну схему:

$$\begin{aligned}\theta_{Enc} &:= \theta_{Enc} - \nabla_{\theta_{Enc}} \left( \mathcal{L}_{prior} + \mathcal{L}_{llike}^{Disl} \right) \\ \theta_{Dec} &:= \theta_{Dec} - \nabla_{\theta_{Dec}} \left( \gamma \mathcal{L}_{llike}^{Disl} - \mathcal{L}_{GAN} \right) \\ \theta_{Dis} &:= \theta_{Dis} - \nabla_{\theta_{Dis}} \left( \mathcal{L}_{GAN} \right)\end{aligned}$$

де  $\gamma$  – коефіцієнт, що регулює важливість функцій помилки ( $\mathcal{L}_{llike}^{Disl}$  – контент, а  $\mathcal{L}_{GAN}$  – стиль).

# Результати генерації змішаної моделі





# Висновки

- Оглянуто теоретичні відомості з даної тематики, а саме:
  - математичне обґрунтування згорткових нейронних мереж;
  - ідея генерації зображення з наперед заданого розподілу (ВАК);
  - метод створення зображень починаючи з сліпого вгадування і закінчуючи стабільною генеративною моделлю (ГДМ);
  - перетворення існуючих зображень в подібні зі зміною прихованих змінних (стилізація).
- Друга та третя частини роботи були присвячені реалізації, тренуванню та аналізу отриманих результатів. Розглянуто процес та кінцевий вид процесу стилізації існуючих зображень ВАК та ГДМ з попередньо тренуваним декодувальником ВАК.
- На практиці було розроблено програмний продукт для тренування та застосування кожної з розглянутих моделей з можливістю зміни вхідних гіперпараметрів.

# Подальші дослідження

Дана робота має потенціал до подальшого дослідження з залученням більшої кількості даних, кардинальною зміною типу зображень (від примітивних монохромних до кольорових зображень живої природи) та розгляду нових архітектур, активаційних функцій, функцій помилки, порядку навчання, тощо. Ідея змагання між мережами вважається найперспективнішою на сьогодні, але їй потребує значних обчислювальних потужностей, що не завжди є загальнодоступними.



Дякую за увагу !