SBER AI

Creative AI

11.09.2020

SBER AI

Денис ДИМИТРОВ

Ph.D. student, Исполнительный директор по анализу данных,
SberAI, Сбер

Картинки были нарисованы человеком или сгенерированы AI?

Смайлик пингвина в синей шляпе, красных перчатках, зеленой рубашке и желтых штанах

Ночь под сиреневой луной

Музыка была написана человеком или сгенерирована AI?

Soul Music

Blues

Генерация дизайна одежды: Acne Studios X Robbie Barrat

Коллекция была показана на **Paris Fashion Week 2020:** collection

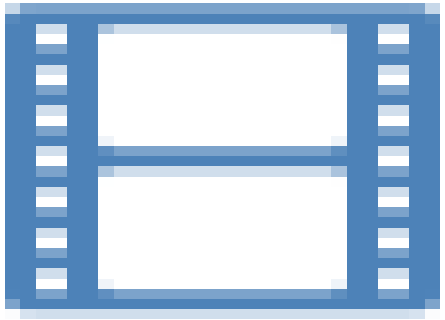
Генерация дизайна одежды: Acne Studios X Robbie Barrat

Thousands of looks from the Acne archives were fed into software by Barrat; Acne creative director Jonny Johansson and his team then created the final looks inspired by the AI renderings.

Acne Studios Men’s Fall Winter 2020 collection in collaboration with Robbie Barrat.

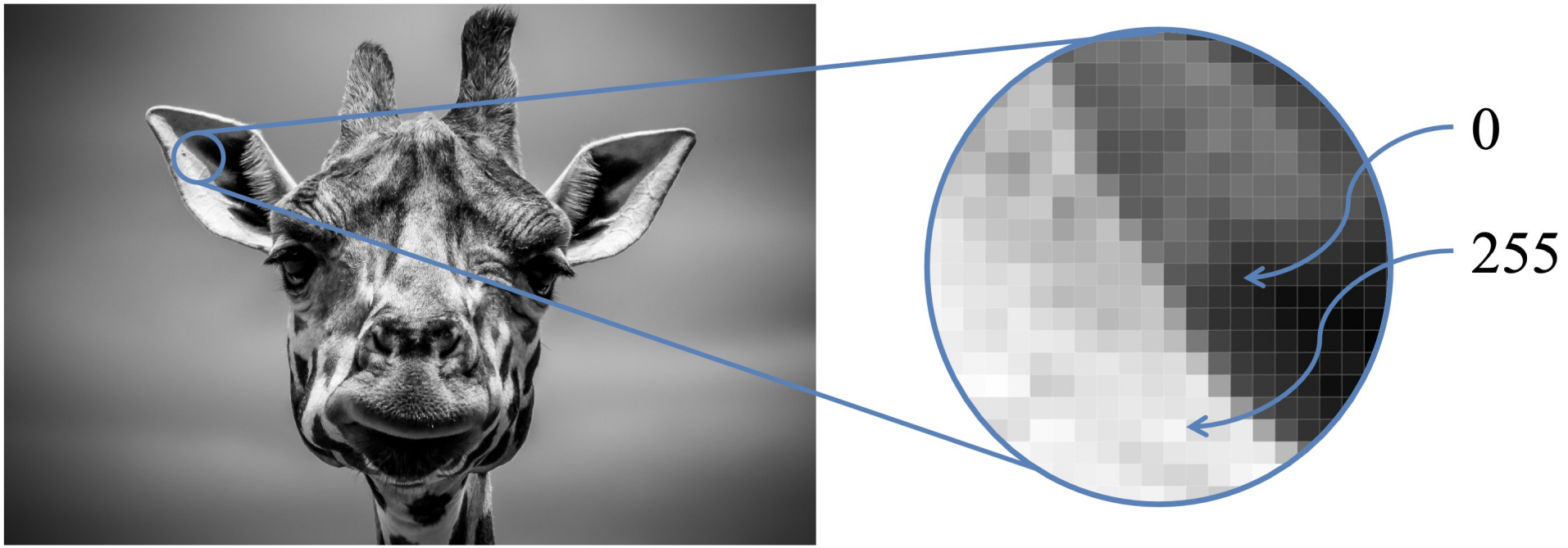
Генерация дизайна одежды: IBM, Tommy Hilfiger and FIT

Генерация принта: Cross & Freckle

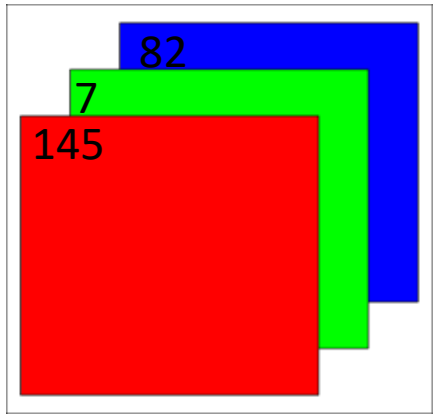


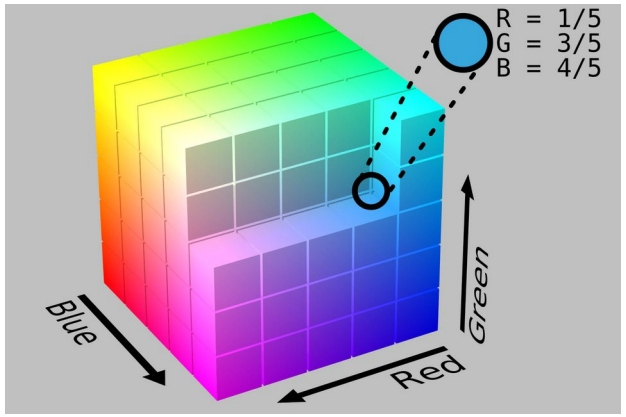
Три модели

sberai.vispstudio.ru


0
255

SBERBANK
UNIVERSITY


82
7
145


R = 1/5
G = 3/5
B = 4/5
Blue
Red
Green

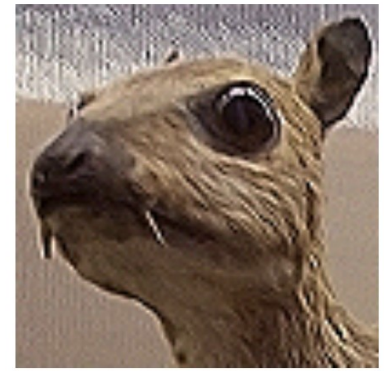
Original image

Kernel

$$* \begin{bmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{bmatrix} =$$

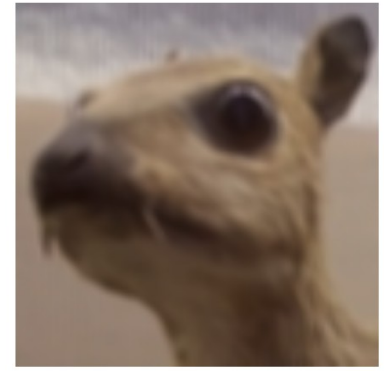
Edge detection

$$* \begin{bmatrix} 0 & -1 & 0 \\ -1 & 5 & -1 \\ 0 & -1 & 0 \end{bmatrix} =$$



Sharpening

$$* \frac{1}{9} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} =$$



Blurring

Neural Style Transfer

«Ваза с 5 подсолнухами», Ван Гог

+

Граффити, Берлин

=

Нидерландский художник-постимпрессионист, не получивший почти никакого специального образования, но за недолгие 10 лет своей творческой карьеры написавший огромное количество полотен, многие из которых стали признанными мировыми шедеврами живописи.

Предыстория

Исследования

В **2015** вышла статья L.A.Gatys, A.S.Ecker and M.Bethge: **A Neural Algorithm of Artistic Style**

Затем последовала серия статей ведущих исследовательских лабораторий мира по улучшению предложенного метода (OpenAI, Google AI) и open source реализаций.

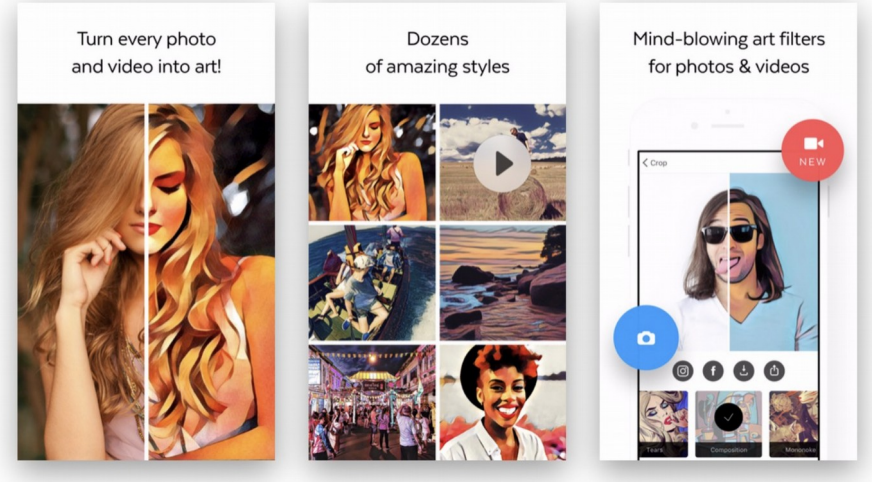
Некоторые из них:

L.A.Gatys, A.S.Ecker and M.Bethge: **Controlling Perceptual Factors in Neural Style Transfer**, 2017
V. Dumoulin, J.S. Shlens, M. Kudlur: **A Learned Representation for Artistic Style**, 2017

Индустрия

DEEPART.io

PRISMA LABS



Artisto

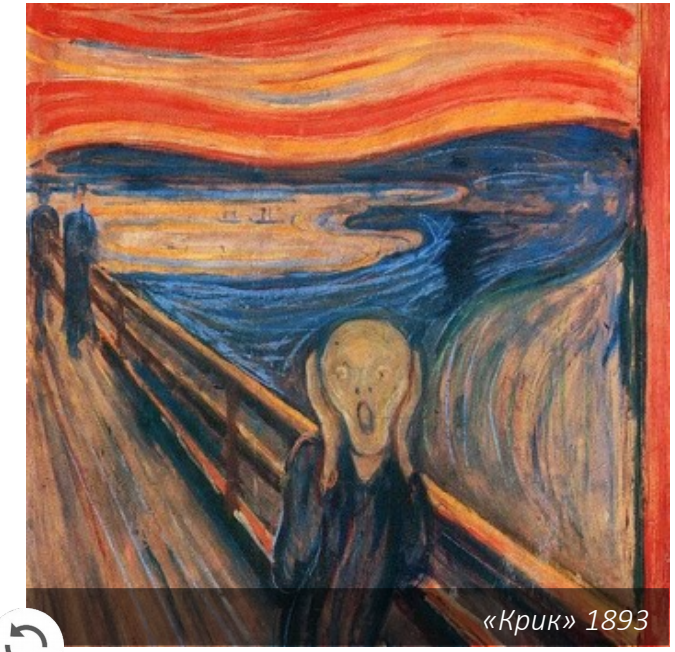
Клод Моне

«The Japanese Bridge» 1899. («The Water-Lily Pond»)

Bicileta Sem Freio, Лос Анджелес

Клод Моне стал одним из основоположников нового направления в живописи - импрессионизма. Главная цель художников-импрессионистов - почувствовать красоту момента и выразить это на полотне. Они писали крупными мазками, использовали чистые цвета, отказываясь от привычного смешивания.

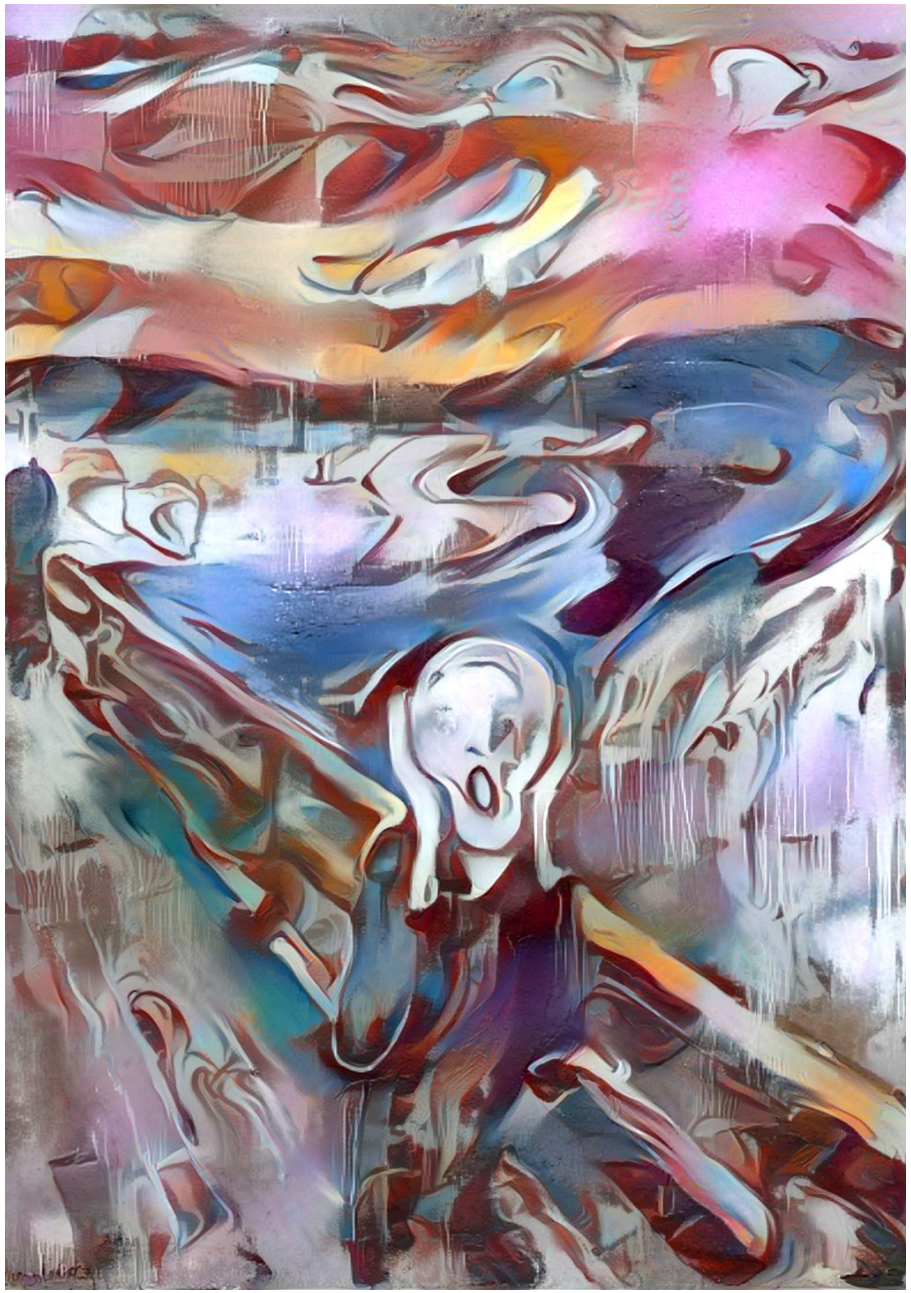
Эдвард Мунк



«Крик» 1893

Граффити, Италия

Норвежский художник с тревожным, северным, обреченным, пророческим мироощущением, которое и легло в основу его живописного стиля – экспрессионизма.

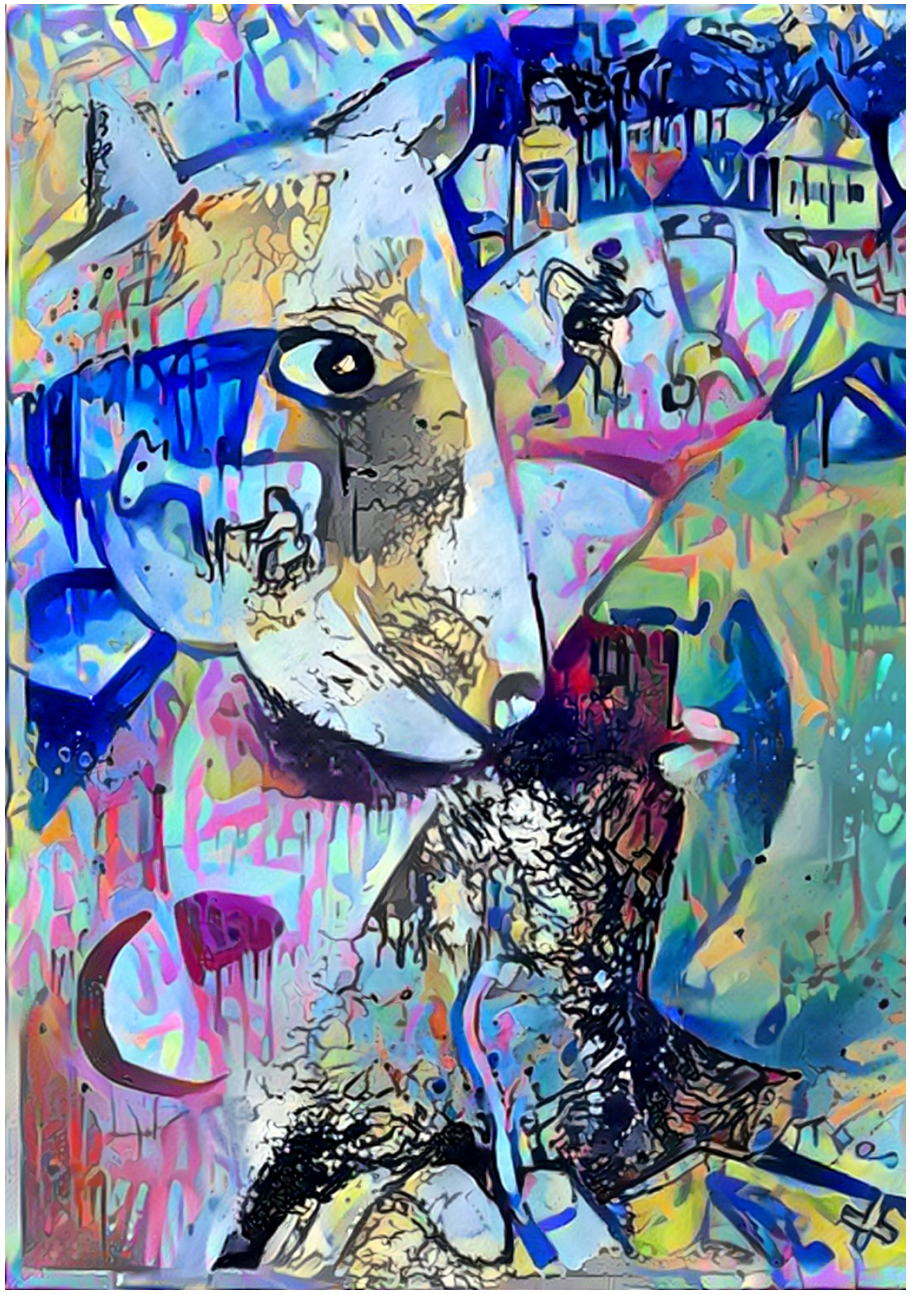


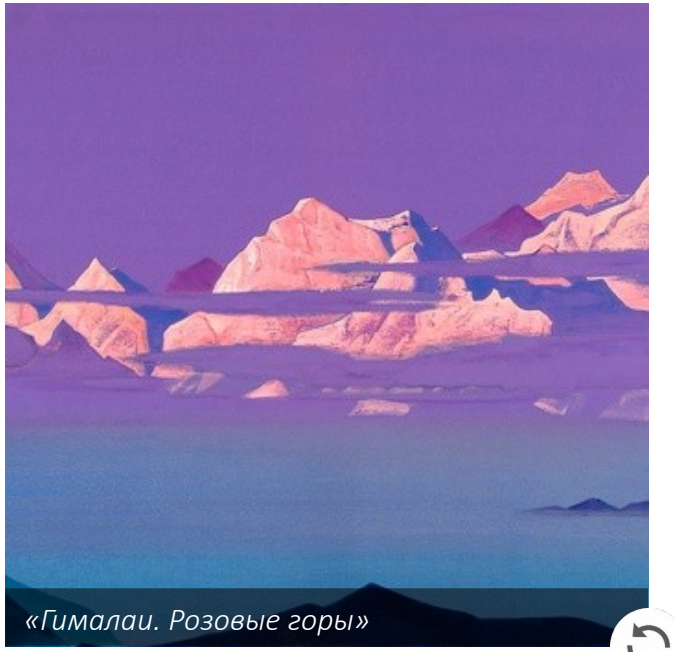
«Я и моя деревня» 1911

Марк Шагал

Еврейский художник, который жил в России, Франции, Америке и всё же остался еврейским художником, сохранил свою самобытность. Шагал создал собственный уникальный стиль в живописи, а также делал фрески и витражи, разрабатывал театральные костюмы и декорации, иллюстрировал книги и сам писал их.

Alec Monopoly, “Godfather”





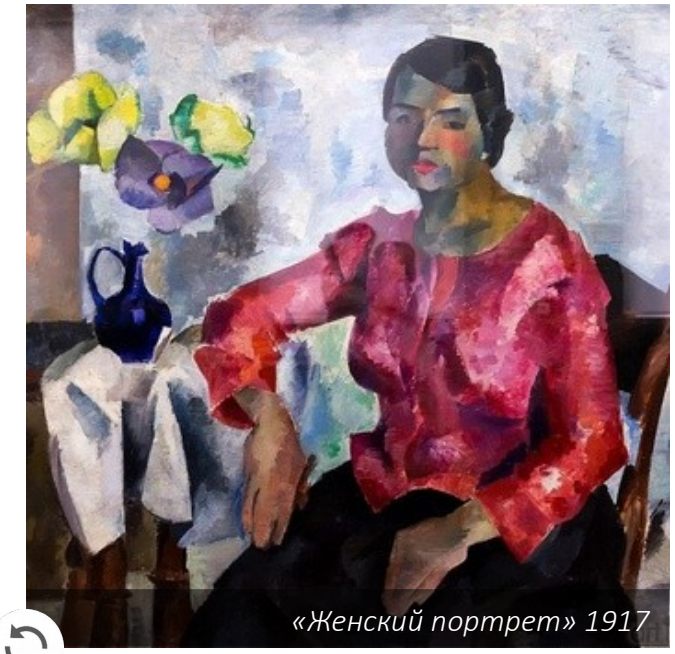
«Гималаи. Розовые горы»

Николай Константинович Рерих

Русский художник, философ, писатель, сценограф, археолог, путешественник. Автор «Пакта Рериха», на законодательном уровне закрепившего охрану и защиту памятников культуры во всем мире.

Граффити, Мельбурн

Роберт Фальк



«Женский портрет» 1917

Автор неизвестен, Нью-Йорк

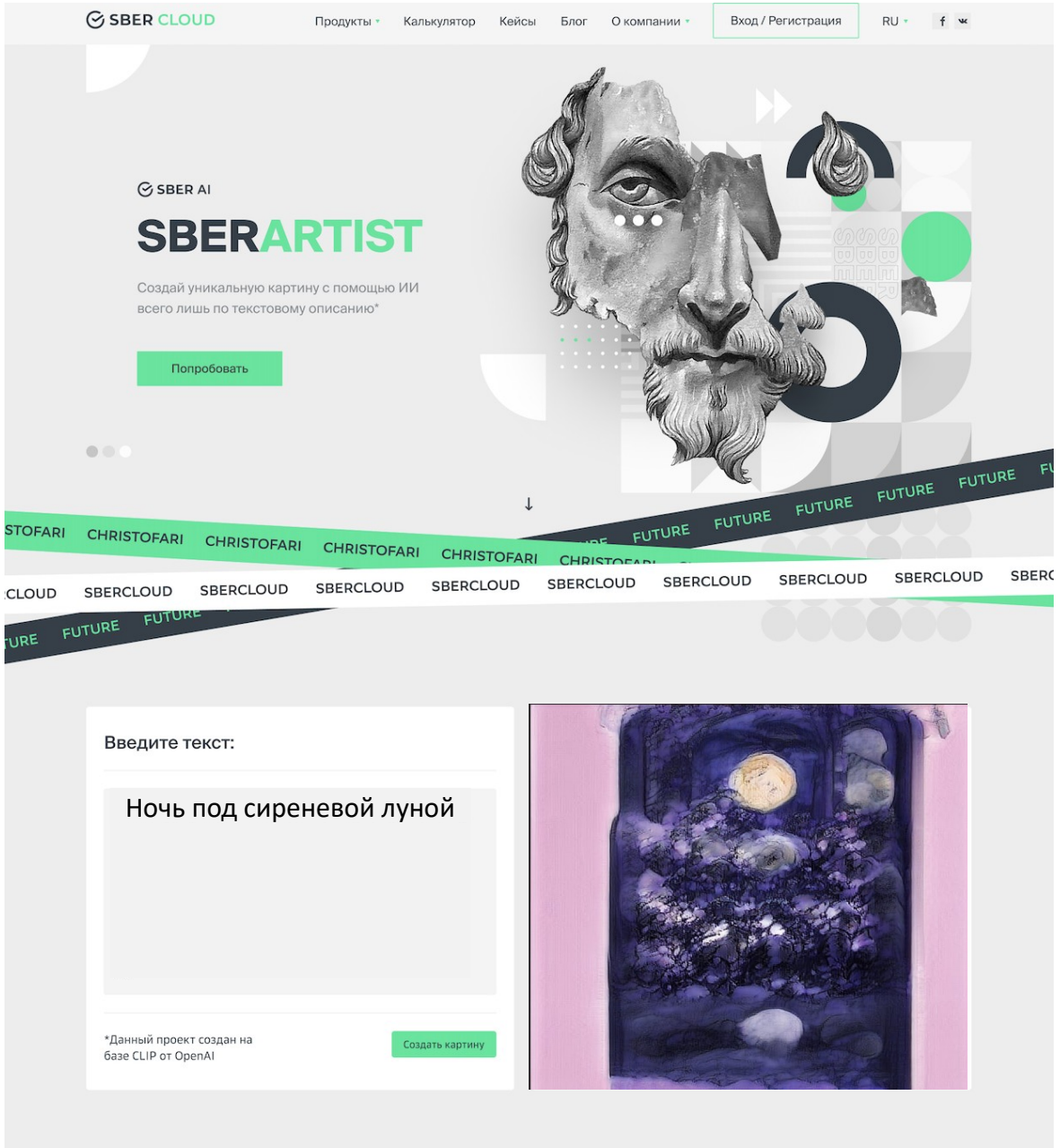
- Роберт Фальк состоял в авангардистском объединении «Бубновый валет» и писал картины в самых разных жанрах: натюрморты, портреты, пейзажи. В 1950-х годах самый «тихий» и лиричный из русских авангардистов стал для последующих поколений патриархом московской школы XX века.

SBER AI

Модель генерирует картинку по текстовому запросу

Использует эмбеддинги CLIP-ViT от OpenAI с генерацией на основе BigGAN.

CLIP-ViT была обучена на датасете объемом 400 млн пар картинка и соответствующее ей текстовое описание. Умеет находить для произвольной картинки лучшее описание и, наоборот, для любого описания подбирать ближайшую по смыслу картинку.



DALL-E

SBER AI

https://openai.com/blog/dall-e/

TEXT PROMPT

an armchair in the shape of an avocado [...]

AI-GENERATED IMAGES

Edit prompt or view more images ↓

DALL-E

SBER AI

TEXT PROMPT a professional high quality emoji of a lovestruck cup of boba

AI-GENERATED IMAGES

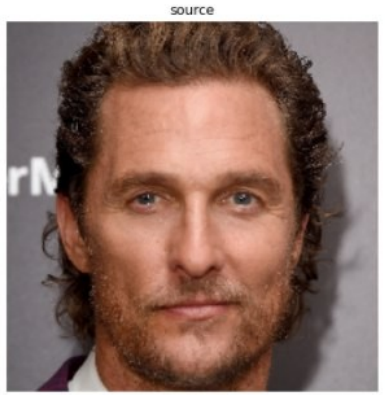
One-shot FaceSwap

Модель умеет переносить произвольное лицо человека на заданное фото или видео (причем достаточно одной фотографии человека)

Модель обучена на двух датасетах лиц: VoxCeleb и LFW Face Database, суммарный размер 1 млн фото лиц людей в разных ракурсах.

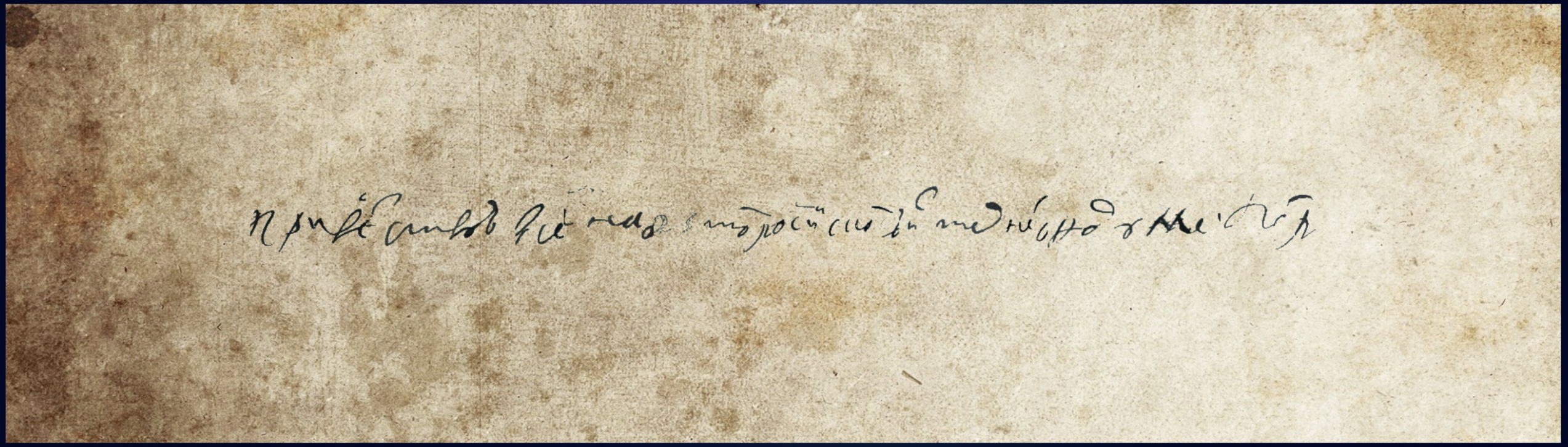
Сложная архитектура на основе GAN.

sberai.vispstudio.ru



Генерация почерка

SBER AI



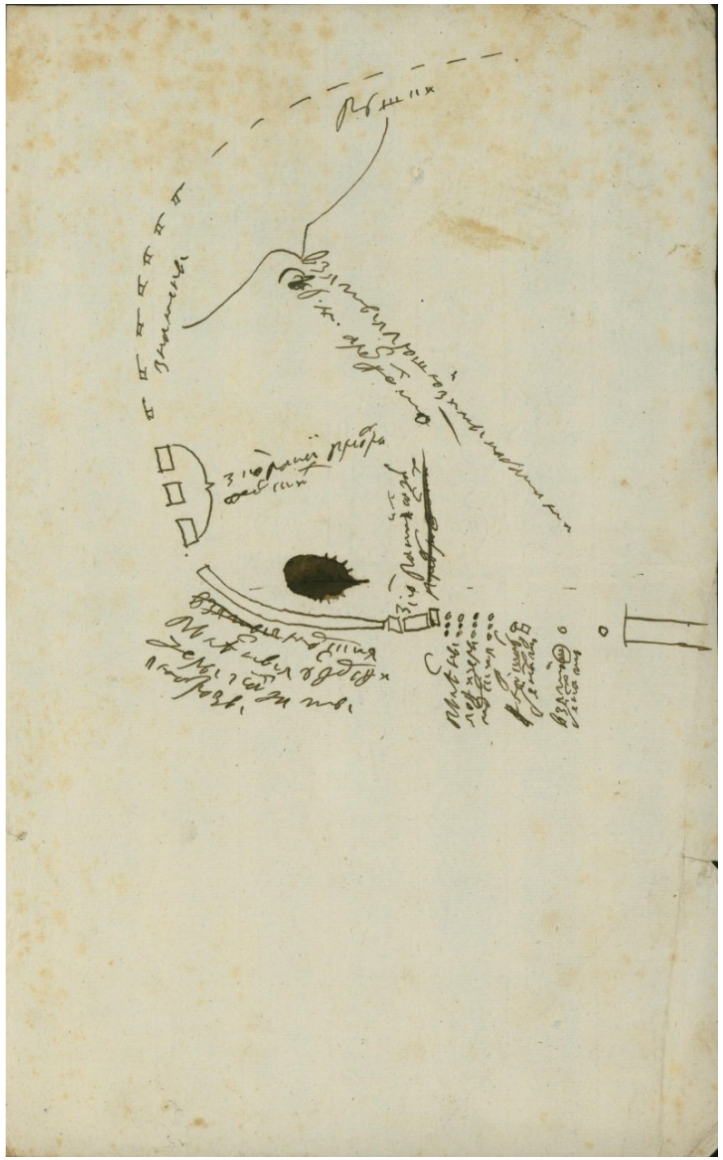
Напишите слово или фразу

Приветствую всех на 25-том Российском Интерн…

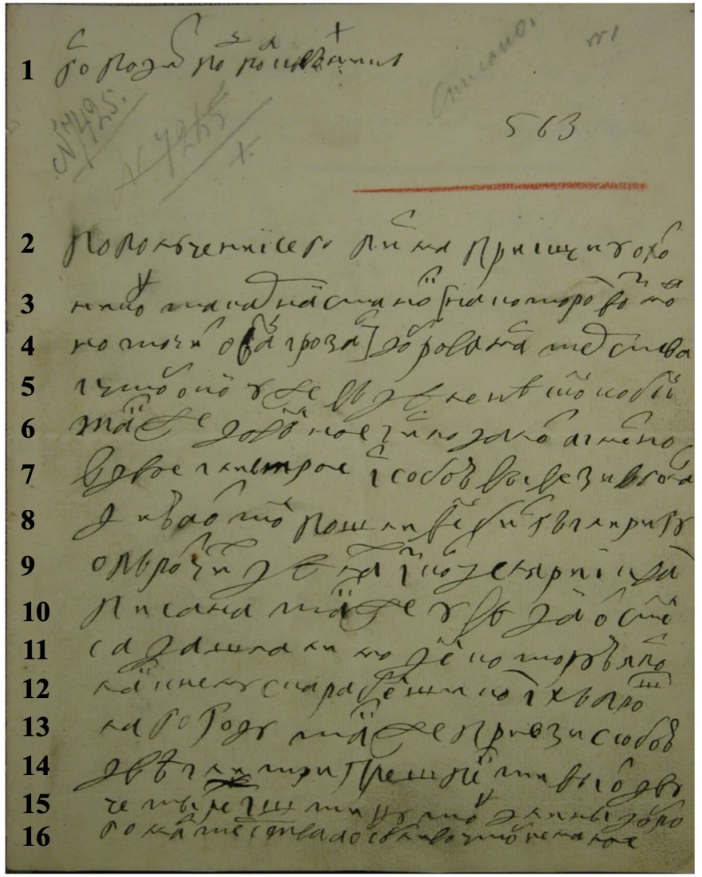
Digital Peter

SBER AI

https://api.sbersight.ru/demo/petr



План парада



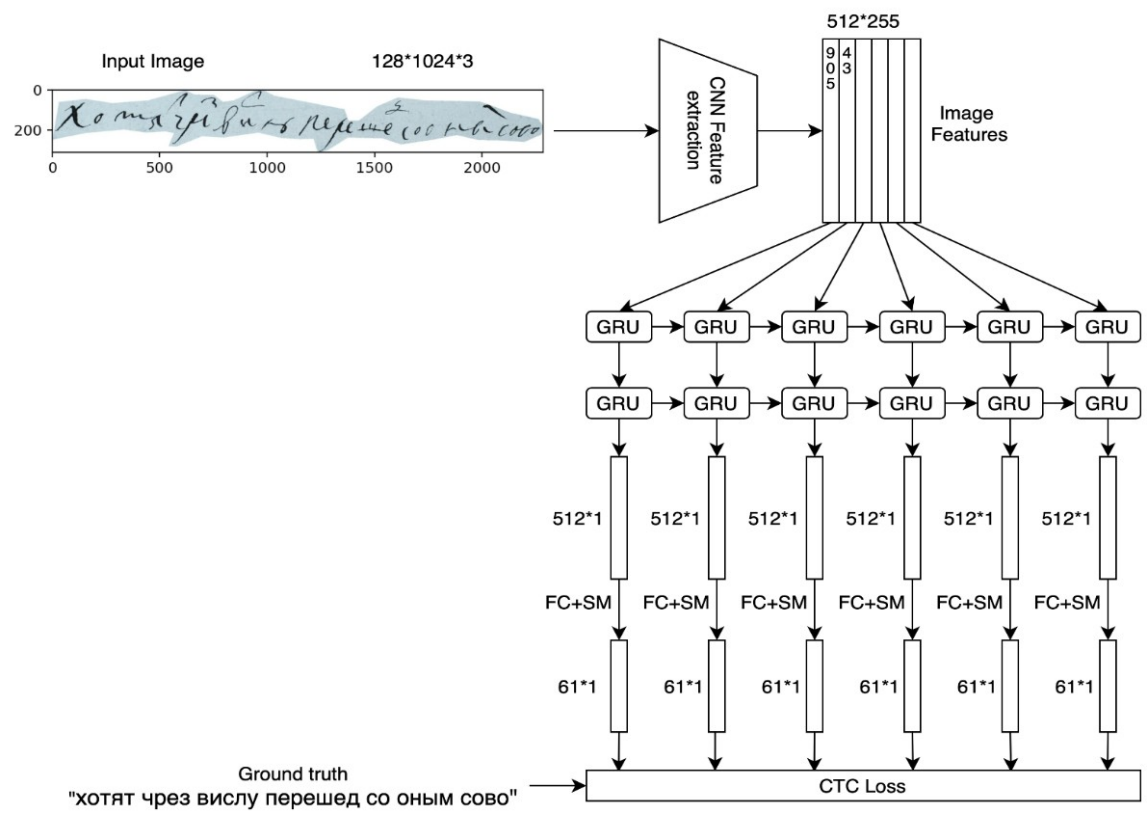
1 Господин подполковъникъ.

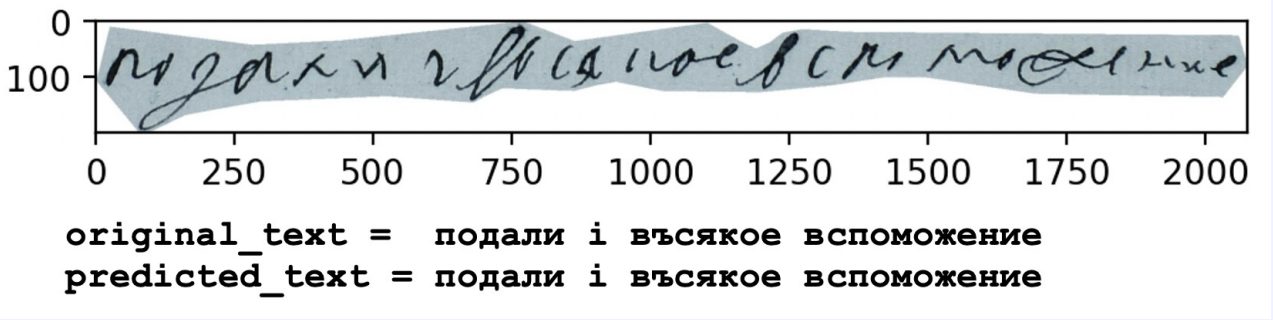
2 По полученіі сего писма приіщи у охот-
3 никоф такарнай станок [на котором возмож-
4 но точит овал і розан] доброва мастерства,
5 і чтоб оной уже въ дѣле нѣсколко был,
6 также доволное число далот, а іменно,
7 вдвое іли втрое, і с собою вывези в Голан-
8 дію, а оттол пошли в Елбингъ іли Ригу.
9 О пъротчих дѣлах із концелярііі к вам
10 писана. Также увѣдай от Стел-
11 са, дашла ли модел, которую я пос-
12 лал к нему с карабелшиком іхъ прош-
13 лаго году. Также прив(е)зи с собою
14 двѣ іли три прешпективы от дву,
15 четырех і шти футоф длины добро-
16 го мастерства, а особливо чтоб немалое

Письмо

Digital Peter

SBER AI





97.6%

Recognition accuracy*

Competition

***1-CharacterErrorRate**
Can be interpreted how the quantity is correct recognized characters

The best model that recognizes the handwriting of Peter I for 2020

Analogue on the market

Recognition accuracy 1-CharacterErrorRate = 68.1%
Model for recognition of cursive writing in Transkribus

Data

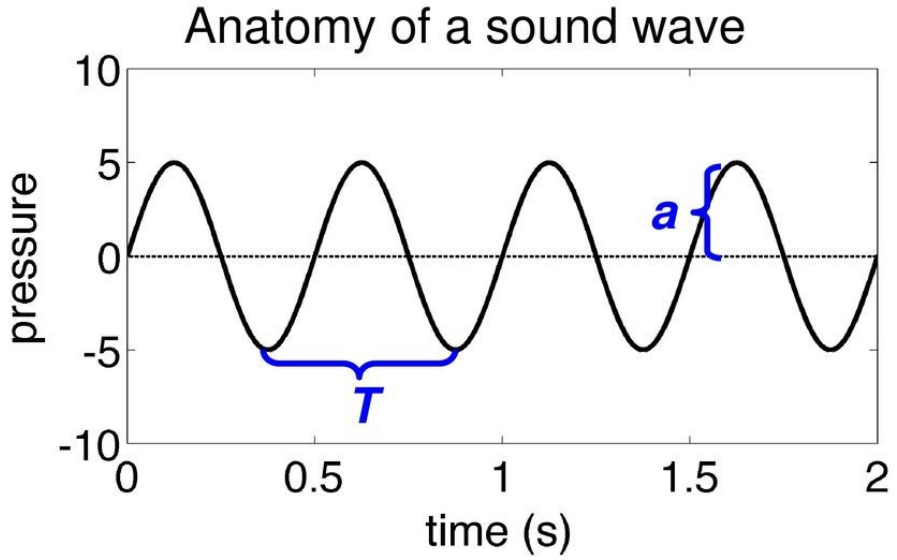
10,058 lines written by the hand of Peter the Great (early 18th century)

Model

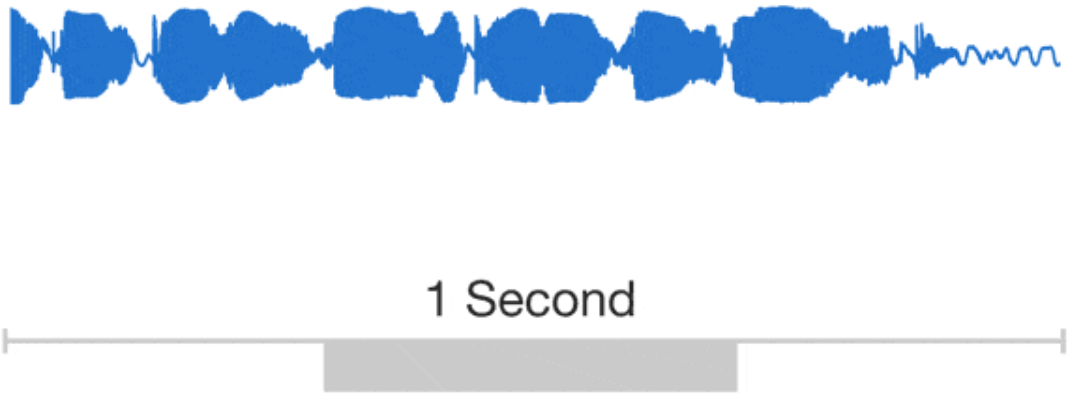
Architecture CRNN + CTC Loss, Shi B., et al.
Also, for spell-checking, the language Transformers based model trained on a dataset of 17th century texts

Как кодируется музыка?

Звук – это волна с непрерывно меняющейся частотой и амплитудой



a = amplitude of wave
T = period of wave
f = frequency = $1/T$
λ = wavelength (= $cT = c/f$, where c is sound speed)



Генеративная музыка

Raw audio 44.1k samples per second, where each sample is a float that represents the amplitude of sound at that moment in time

Encode using CNNs
(convolutional neural networks)

Compressed audio 344 samples per second, where each sample is 1 of 2048 possible vocab tokens

Generate novel patterns from trained transformer conditioned on lyrics

Novel compressed audio 344 samples per second

Upsample using transformers
and decode using CNNs

Novel raw audio 44.1k samples per second

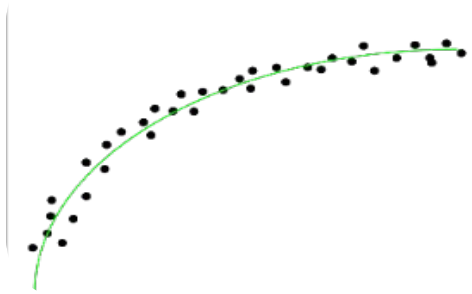
https://go.zvooq.com/birth

Спасибо за внимание!

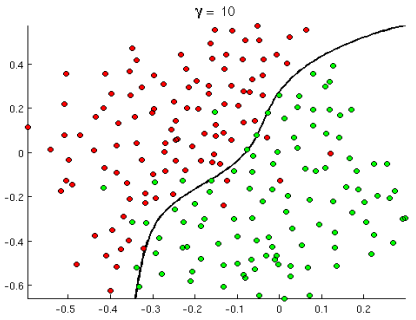
Машинное обучение

— про то, как восстановить сложные закономерности и зависимости по конечному числу примеров (размеченных данных)

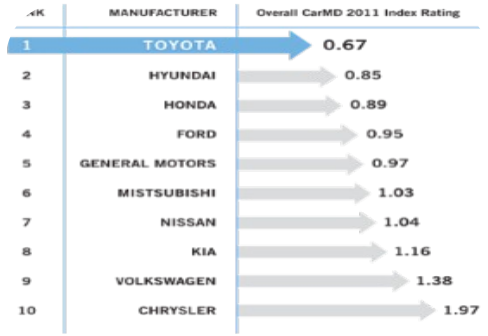
Мы будем в основном рассматривать обучение с учителем: **аппроксимировать** $y(x)$, если известна **обучающая выборка**: $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$



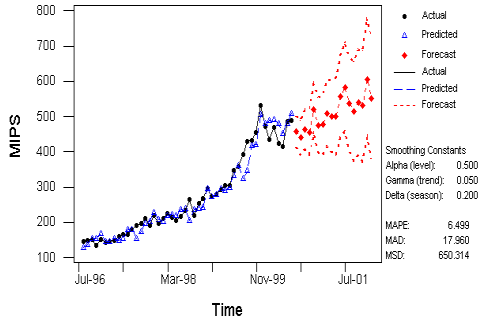
Регрессия



Классификация



Ранжирование



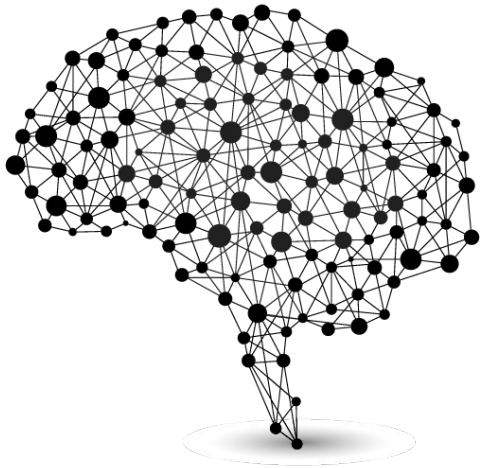
Прогнозирование

Общая постановка

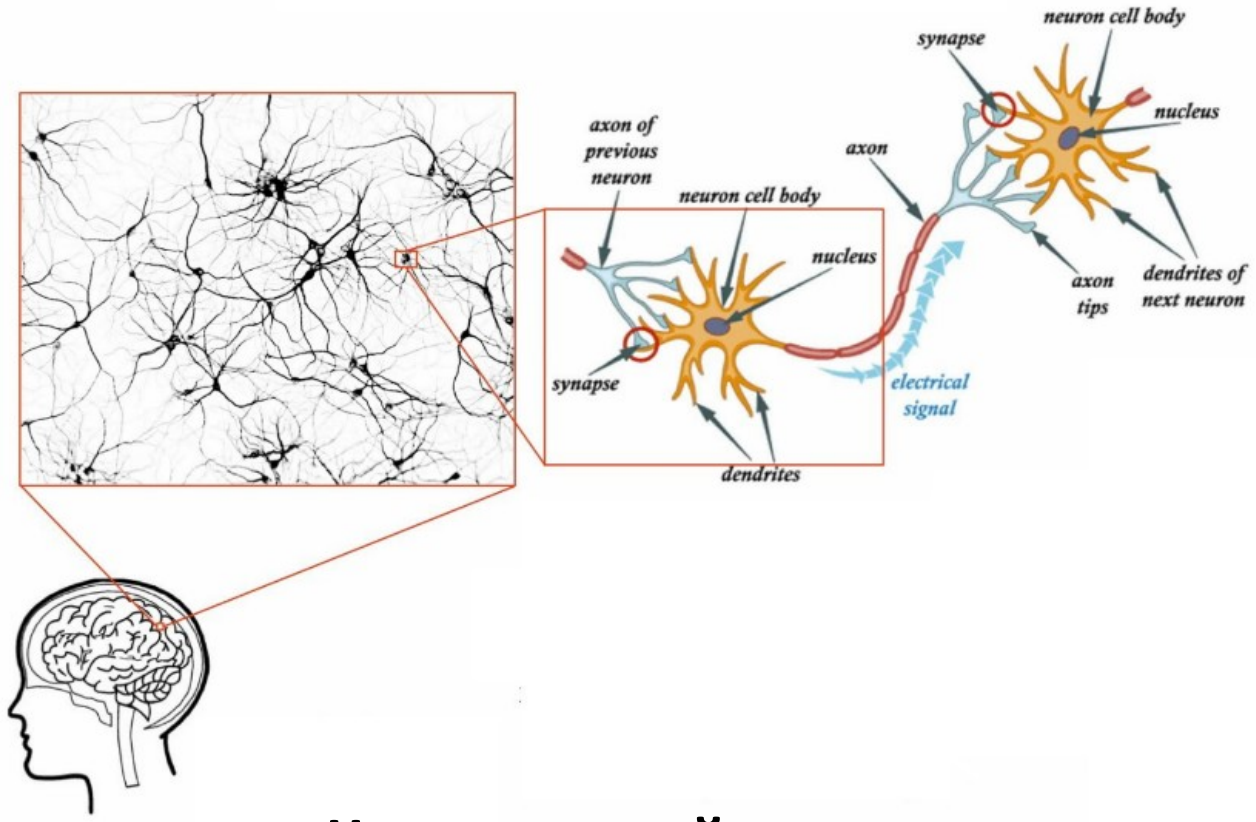
$$x \xrightarrow{f(x)} y$$

$$y \approx f(x)$$

Есть очень много способов эту зависимость восстановить (т.е. отыскать функцию f). Мы рассмотрим нейронные сети (глубокое обучение).

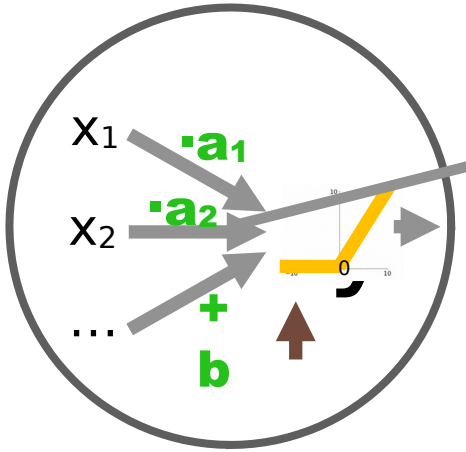


Нейронные сети

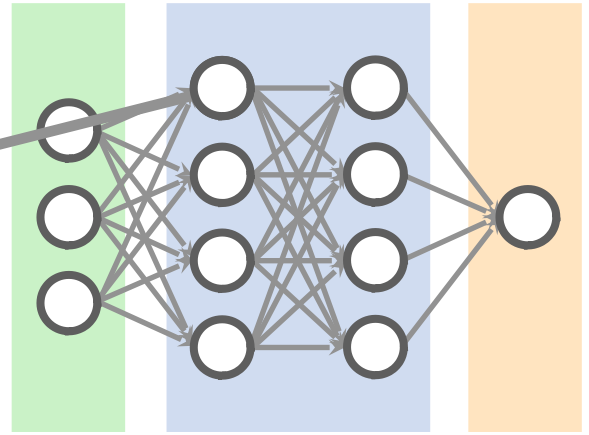


Человеческий мозг

и



Модель простейшего нейрона

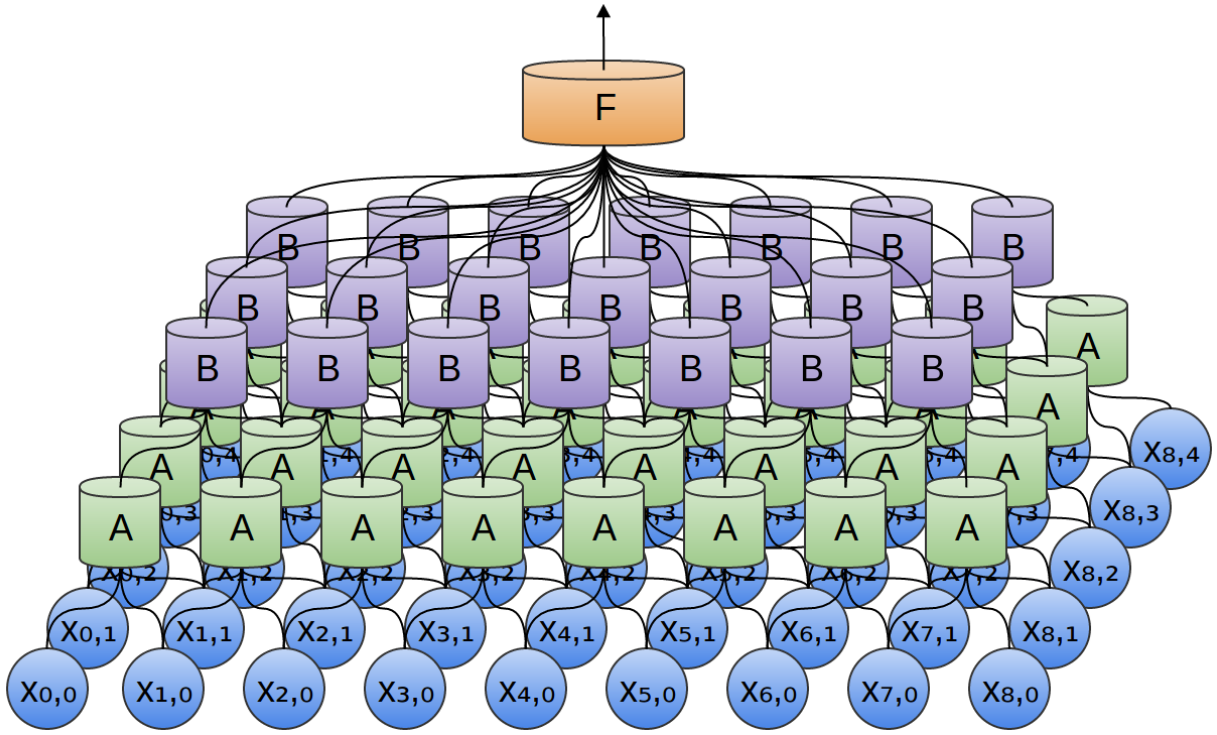


x Скрытые слои с нейронами y

Искусственная нейронная сеть - связанная группа узлов, схожая с огромной сетью нейронов в мозге. Здесь каждый узел представляет собой

связь выхода одного нейрона со входом другого.

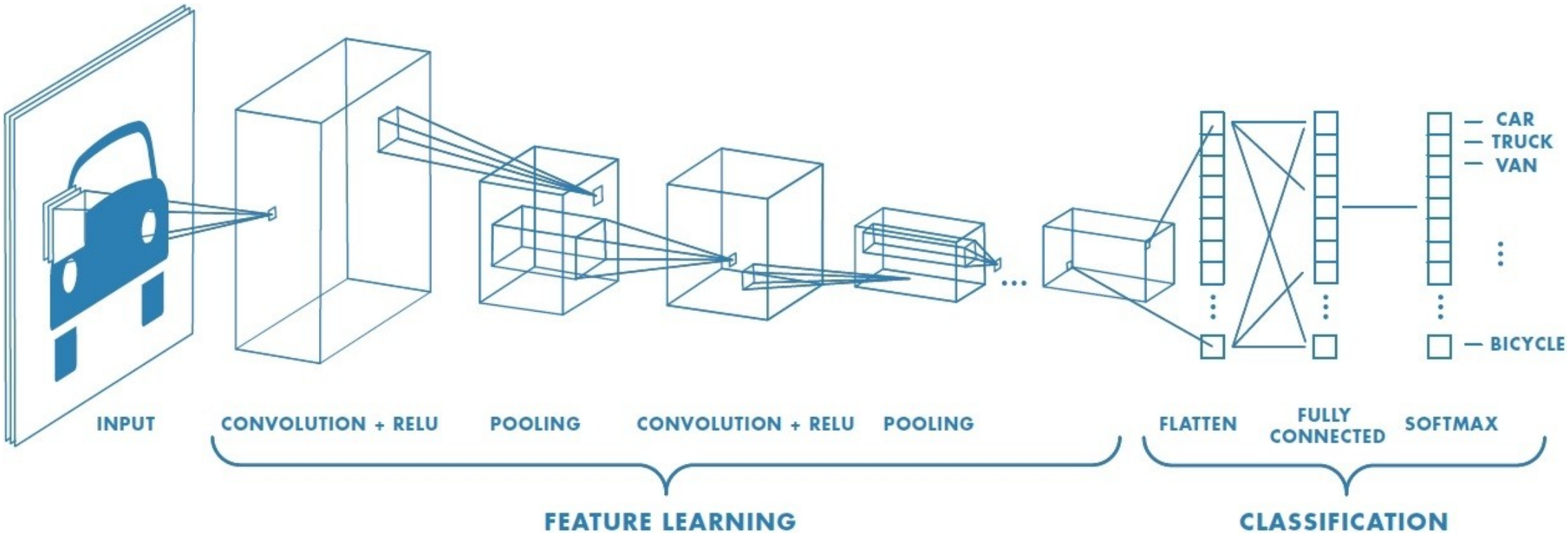
Свёрточные нейронные сети (CNN)



Двумерная свёрточная нейронная сеть

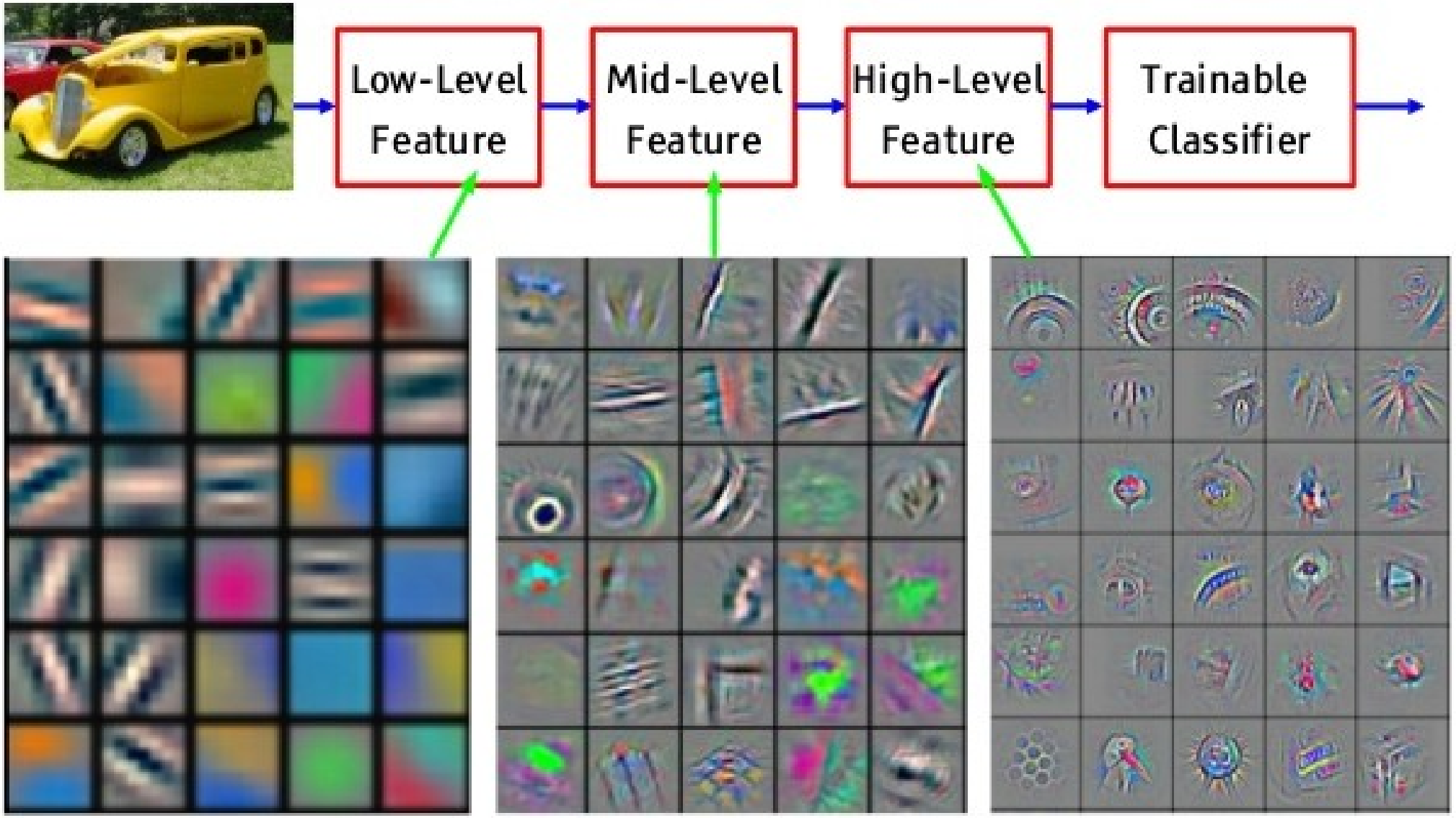
- Идея основана на структуре зрительной коры животных
- A, B, … - свёрточные слои, каждый слой вычисляет свой **признак**

Классификация изображений (CNN)

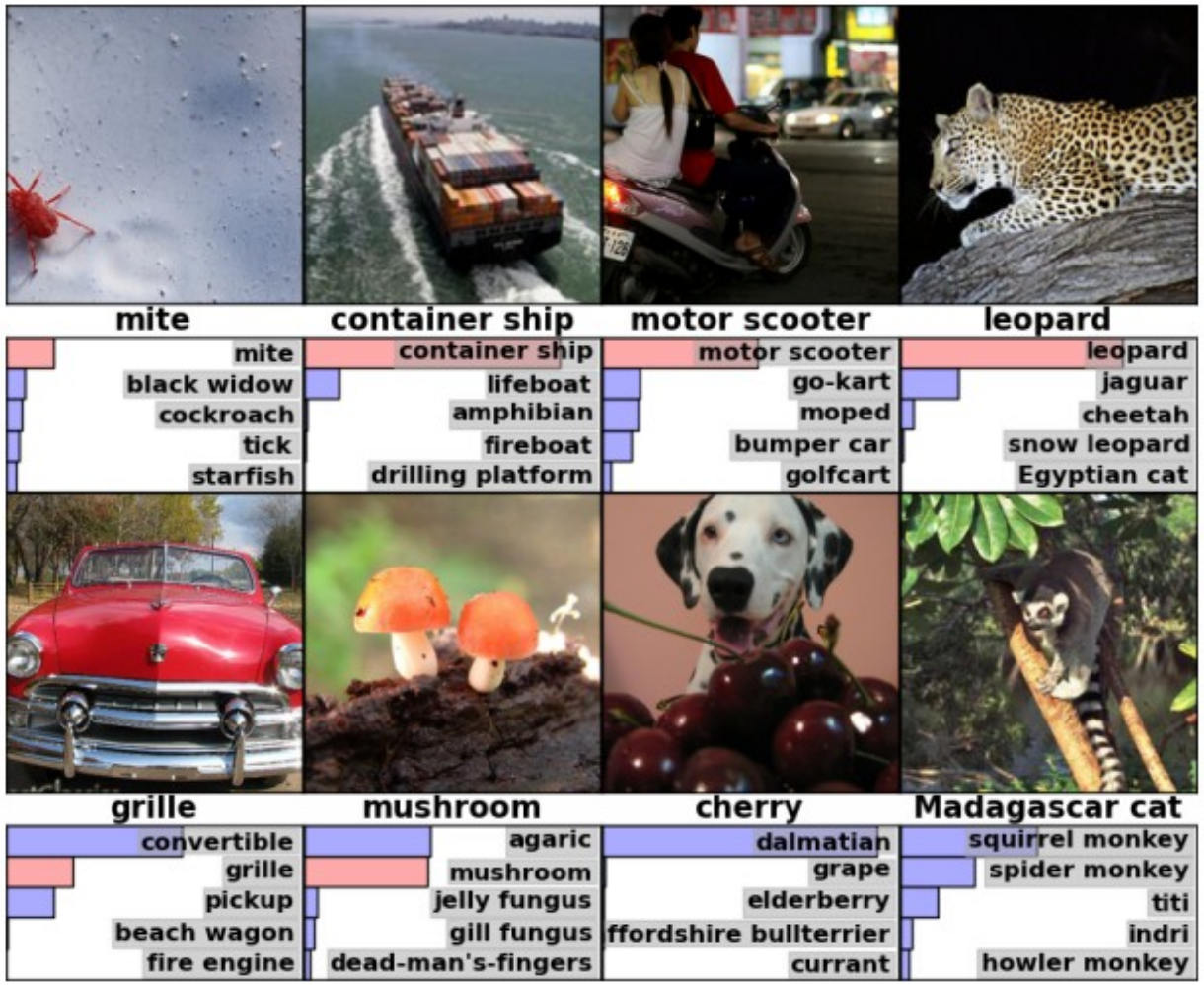


- Yan LeCun (CNN, 1988)
- Обучение нейронной сети классифицировать изображения в тысячи различных классов
- **ImageNet** – очень большой набор изображений различных классов

Свёрточные нейронные сети



Классификация изображений

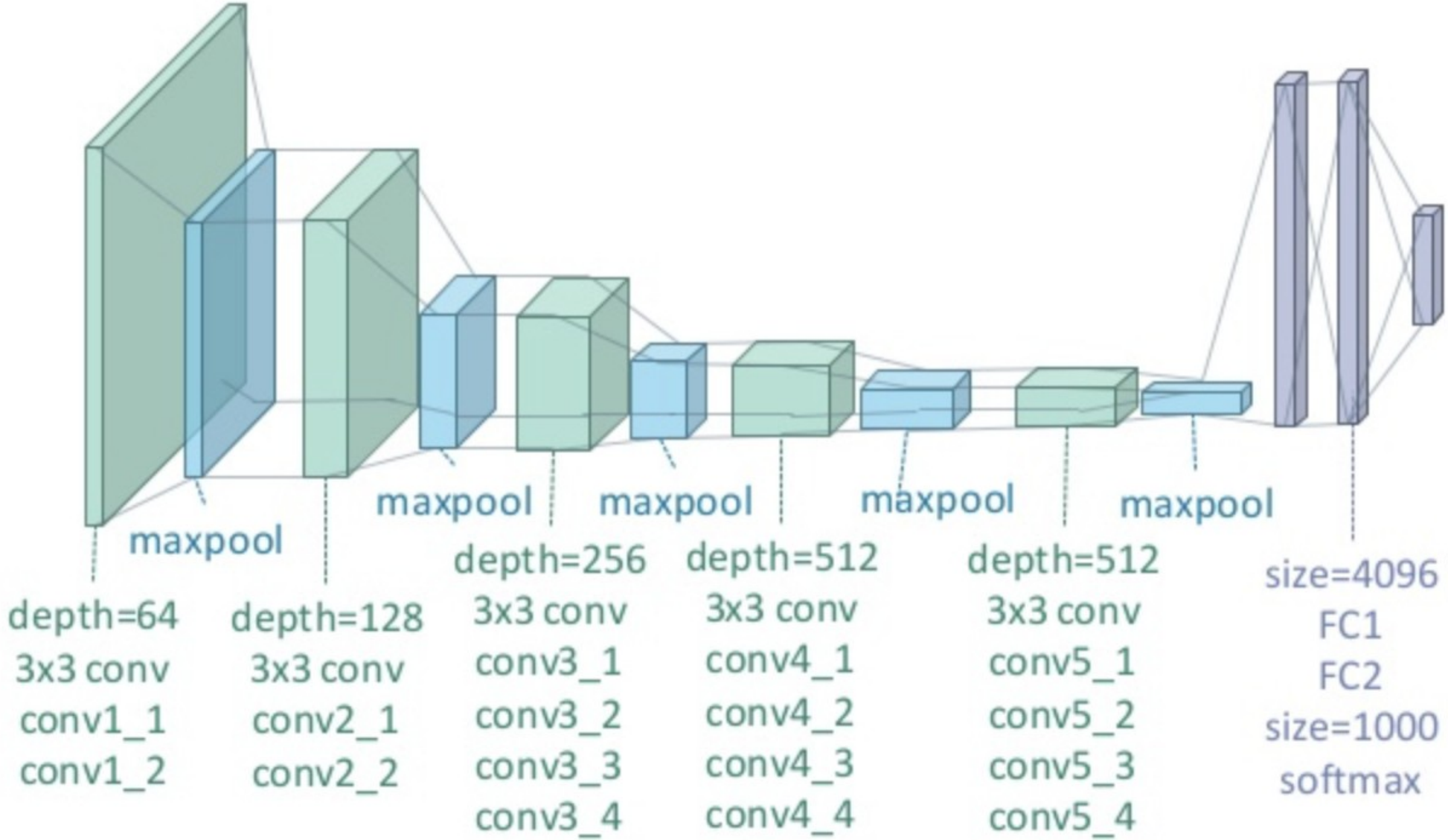


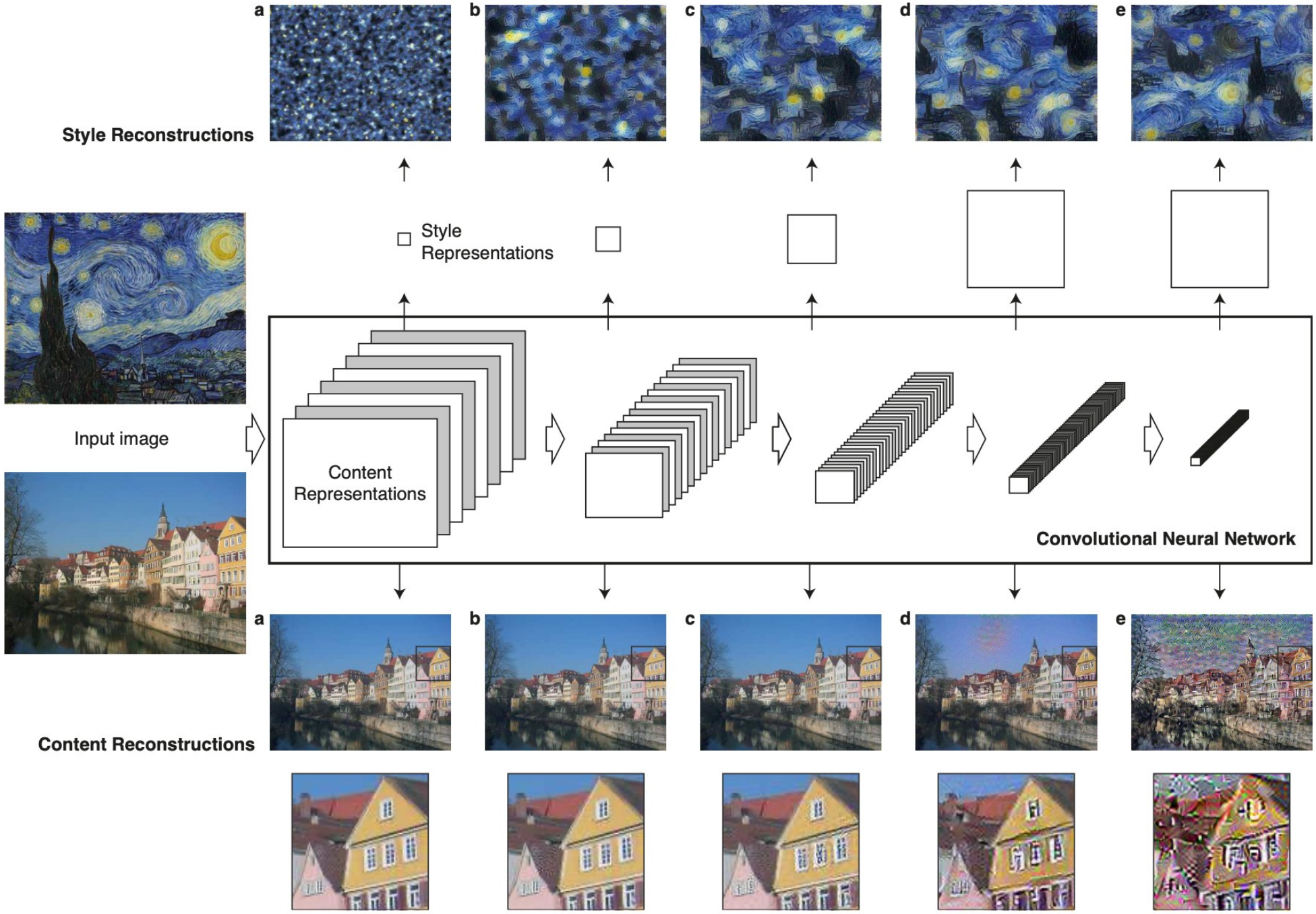
Сверху: 4 верно классифицированных изображения. Снизу: 4 неверно

к

наиболее вероятных классов, предсказанных моделью. Модель дает правильный ответ в 63% случаев, в 85% случаев ответ содержится среди 5 наиболее вероятных классов.

VGG-19


maxpool
maxpool
maxpool
maxpool
maxpool
depth=64
3x3 conv
conv1_1
conv1_2
depth=128
3x3 conv
conv2_1
conv2_2
depth=256
3x3 conv
conv3_1
conv3_2
conv3_3
conv3_4
depth=512
3x3 conv
conv4_1
conv4_2
conv4_3
conv4_4
depth=512
3x3 conv
conv5_1
conv5_2
conv5_3
conv5_4
size=4096
FC1
FC2
size=1000
softmax


a
b
c
d
e
Style Reconstructions
Style
Representations
Input image
Content
Representations
Convolutional Neural Network
a
b
c
d
e
Content Reconstructions

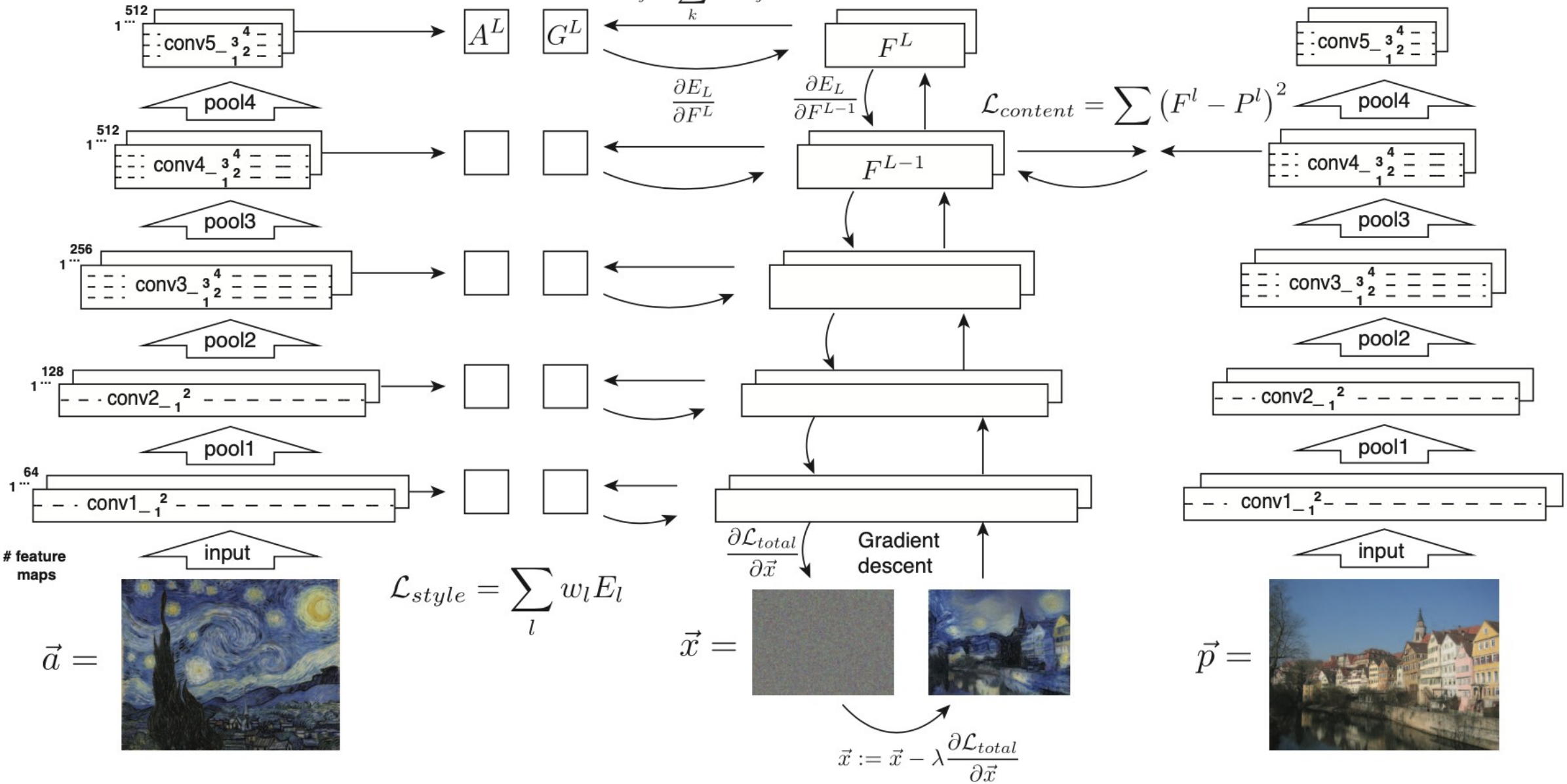
$$E_L = \sum \left(G^L - A^L\right)^2 \qquad \mathcal{L}_{total} = \alpha \mathcal{L}_{content} + \beta \mathcal{L}_{style}$$

$$G^L_{ij} = \sum_k F^L_{ik} F^L_{jk}.$$

$$\mathcal{L}_{content} = \sum \left(F^l - P^l\right)^2$$

$$\mathcal{L}_{style} = \sum_l w_l E_l$$

$$\vec{x} := \vec{x} - \lambda \frac{\partial \mathcal{L}_{total}}{\partial \vec{x}}$$



Контент

$$\mathcal{L}_{\text{content}}(\vec{p}, \vec{x}, l) = \frac{1}{2} \sum_{i,j} \left(F_{ij}^{l} - P_{ij}^{l}\right)^2$$

Стиль

$$G_{ij}^{l} = \sum_{k} F_{ik}^{l} F_{jk}^{l}.$$

$$E_l = \frac{1}{4N_l^2 M_l^2} \sum_{i,j} \left(G_{ij}^{l} - A_{ij}^{l}\right)^2$$

$$\mathcal{L}_{\text{style}}(\vec{a}, \vec{x}) = \sum_{l=0}^{L} w_l E_l,$$

$$\mathcal{L}_{\text{total}}(\vec{p}, \vec{a}, \vec{x}) = \alpha \mathcal{L}_{\text{content}}(\vec{p}, \vec{x}) + \beta \mathcal{L}_{\text{style}}(\vec{a}, \vec{x})$$

Минимизируем эту функцию по x

Controlling Style Transfer*

*L.A.Gatys et all: **Controlling Perceptual Factors in Neural Style Transfer, 2017**.

(a) Content

(b) Style I

(c) Style II

(d) Output using [8]

(e) Output with spatial control

(f) Output spatially combining styles I and II

(a) Content

(b) Style I/II

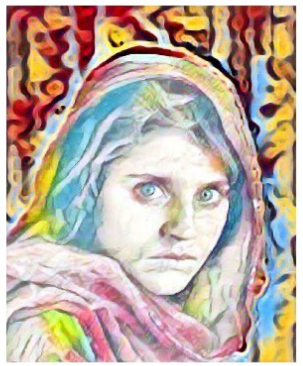
(c) Output with style I/II

(d) Original network

(e) Luminance network

(f) Vertical mask

(g) Horizontal mask


(h) Person/Background mask from [21]

(a) Content (b) Style I/II/III

(c) Style IV/V (new) (d) Output with style II

(e) Output with style IV

(f) Output with style V