

Identyfikacja autorów obrazów

Katsiaryna Yablonskaya

1. Wstęp

Identyfikacja artysty utworu jest trudnym problemem, rozwiązywanym przede wszystkim przez historyków sztuki potrzebujących długoterminowego wykształcenia i specjalistycznej wiedzy. Przyszli eksperci uczą się z podręczników wymieniających cechy artystów, stopniowo rozwijając intuicję. Jest to złożony i interesujący problem dla uczenia maszynowego, ponieważ identyfikacja artystów wymaga nie tylko wykrycia obiektu lub twarzy, ale także identyfikacji stylu i cech ludzkiej kreatywności.

Celem naszego projektu jest wyszkolenie sieci neuronowej, tak aby rozpoznała autorów obrazów tak dokładnie, jak to możliwe. Celem było uzyskanie ponad 50% dokładności dla 10 klas określonych artystów.

Project zrobiony przy użyciu biblioteki keras i frameworku tensorflow.

Zestaw danych składa się ze 220 zdjęć 10 popularnych artystów różnych stylów, krajów i czasów:

Chagall (1887-1985) – Rosyjsko-francusko-żydowski modernista

Dali (1904-1989) – Hiszpański surrealista

Gauguin (1848-1903) – Francuski symbolista, postimpresjonista

Hokusai (1760-1849) – Japoński okres Edo

Malevich (1879-1935) – Ukraińsko-rosyjsko-polski suprematysta

Picasso (1881-1973) – Hiszpański surrealista, kubista

Raphael (1483-1520) – Włoski renesans

Repin (1844-1930) – Rosyjski realizm

Seurat (1859-1891) – Francuski pointillist

Warhol (1928-1987) – Amerykański pop-art

Jak wspomniano wcześniej, identyfikacją artystów zajmują się przede wszystkim ludzie. Zestaw danych Wikiart, składający się z około 150 000 dzieł sztuki wykonanych

przez 2 500 artystów jest oznaczany przez wolontariuszy. Artsy Art Genome Project jest prowadzony przez ekspertów, którzy ręcznie klasyfikują sztukę.

Większość prób zastosowania uczenia maszynowego do tego problemu opiera się na funkcjach, których celem jest identyfikacja cech, które najskuteczniej odróżniają artystów i style, w tym skalo-niezmienne przekształcenie cech (SIFT), histogram zorientowanych gradientów (HOG) i inne. Do identyfikacji z wykorzystaniem tych funkcji wykorzystywane są maszyny SVM razem z innymi metodami klasyfikacji, takimi jak k-nn i hierarchiczna analiza klastrów.

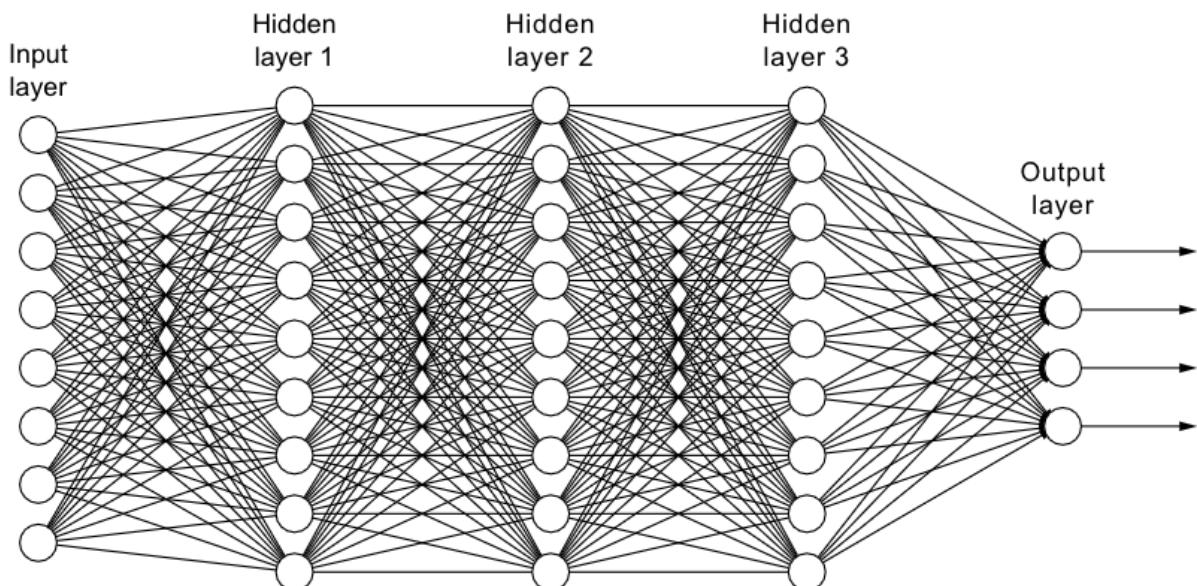
Ostatnio CNN poczynił znaczne postępy w wielu zadaniach związanych z rozpoznawaniem obrazów, takich jak osiągnięcie błędu klasyfikacji 3,6% dla zestawu danych ImageNet. Ta dokładność jest wyższa niż wydajność człowieka w tym samym zbiorze danych.

W zastosowaniu do sztuki, CNN z powodzeniem wykorzystana do rozkładu obrazu na elementy stylu i treści oraz do przenoszenia stylu z jednego obrazu na drugi, co oznacza, że sieć neuronowa jest w stanie uchwycić styl obrazów.

2. Opis metody

Konwolucyjna sieć neuronowa - specjalna architektura sieci neuronowych, zaproponowana przez Yana Lekuna w 1988 r. i mająca na celu skuteczne rozpoznawanie obrazów. Główną ideą konwolucyjnych sieci neuronowych jest przemiana warstw konwolucyjnych (ang. convolution layers) i dokonujących redukcji (ang. pooling layers). Dzięki tej architekturze sieci konwolucyjne szybko się szkolą.

Rozpatrzymy zwykłą sieć neuronową:



Każdy neuron jednej warstwy sieci jest połączony z każdym neuronem następnej.

Jest to bardzo efektywne podejście dla szkolenia sieci, ale taka architektura nie jest efektywną w zadaniu rozpoznawania wzorców, ponieważ nie uwzględnia struktury przestrzennej obrazów. Na przykład, traktuje piksele wejściowe, które są daleko od siebie i blisko siebie, dokładnie na tej samej podstawie.

Aby poprawnie rozpoznać obraz, musimy podkreślić niektóre z jego głównych elementów, co stanowi pewną kombinację, która pozwala nam odróżnić wzorzec. W tym miejscu przydaj się nam **warstwa konwolucyjna**.

Warstwa konwolucyjna

W przeciwieństwie do zwykłych sieci neuronowych, każdy neuron warstwy konwolucyjnej łączymy z kilkoma neuronami (regionem neuronów) poprzedniej warstwy przechodzącym przez pewien filtr ze zbioru filtrów.

Neuron wyjściowy jest iloczynem skalarnym między wpisami filtra a wejściem.

1x1	1x0	1x1	0	0
0x0	1x1	1x0	1	0
0x1	0x0	1x1	1	1
0	0	1	1	0
0	1	1	0	0

4		

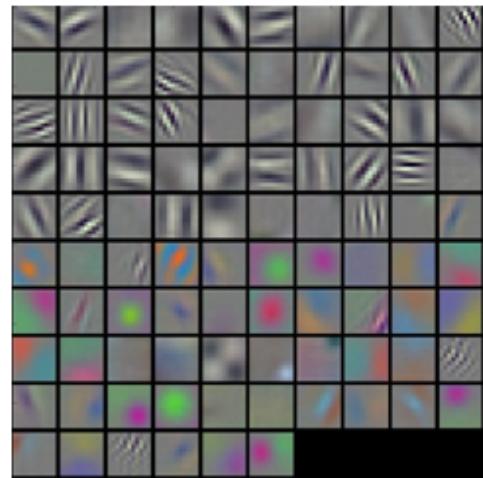
Następnie przesuwamy lokalne pole odbiorcze o jeden piksel w prawo (o jeden neuron), aby połączyć się z drugim neuronem:

1	1x1	1x0	0x1	0
0	1x0	1x1	1x0	0
0	0x1	1x0	1x1	1
0	0	1	1	0
0	1	1	0	0

4	3	

W ten sposób każdy filtr jest użyty na całej szerokości i wysokości tablicy wejściowej, tworząc dwuwymiarową mapę obrazu dla tego filtra. Więc pełna warstwa konwolucyjna składa się z kilku różnych map obiektów.

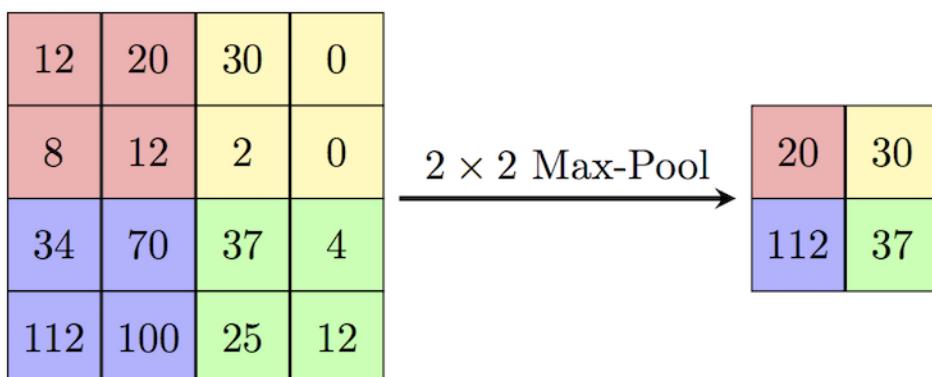
Oprócz właśnie opisanych warstw konwolucyjnych, konwolucyjne sieci neuronowe zawierają również **warstwy dokonującą pool'ing**. Te warstwy są zwykle używane natychmiast po warstwach konwolucyjnych i upraszczają informacje na wyjściu z warstwy konwolucyjnej.



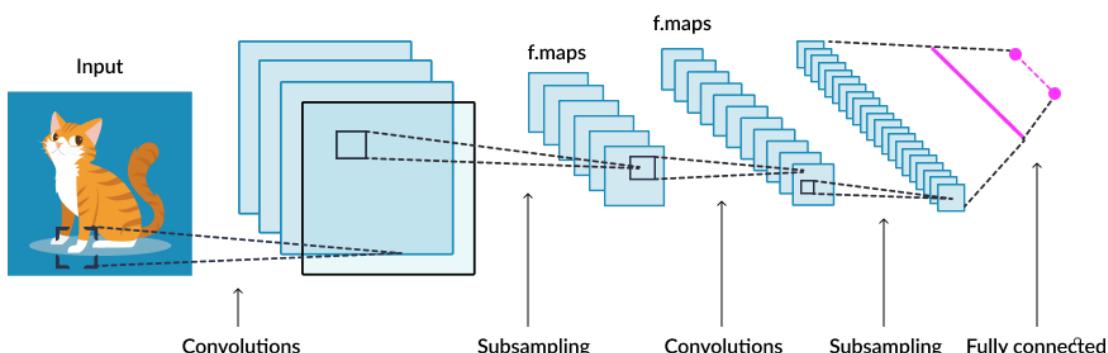
Niekotere filtry uzywane dla utworzenia konwolucyjnej warstwy

Warstwa redukcji (warstwa dokonująca pool'ing)

Ta warstwa pobiera obszar z warstwy konwolucyjnej i przygotowuje mapę skompresowaną. Konkretnym przykładem jest jedna procedura tworzenia tej warstwy, zwana max-pooling. W max-poolingu jednostka warstwy po prostu wysyła maksymalną aktywację w regionie wejściowym 2×2 , jak pokazano na poniższym schemacie:



W ten sposób warstwy mogą się zmieniać na przemian kilka razy aż dojdziemy do ostatniej warstwy - warstwy wyjściowej, która zazwyczaj jest warstwą gęstą (ang. fully connected).



Oznacza to, że ta warstwa łączy każdy neuron z warstwy dokonującej pool'ing do każdego z neuronów wyjściowych.

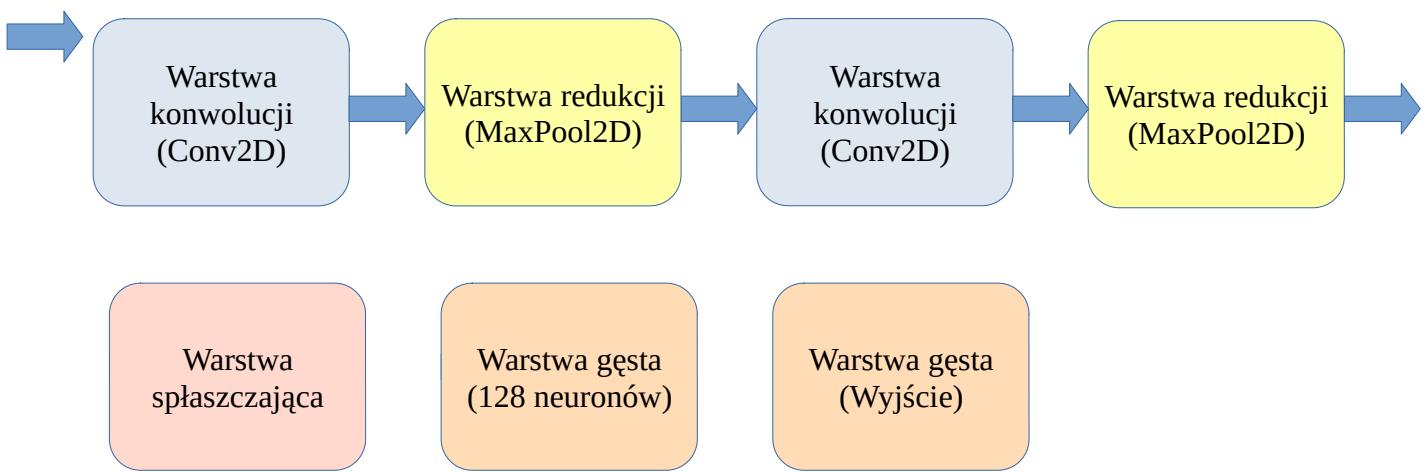
3. Implementacja

Do szkolenia sieci wykorzystaliśmy zestaw danych zawierający 220 obrazów 10 artystów, podzielony na zestaw szkoleniowy (150 obrazów) i zestaw testowy (70 obrazów). Ponadto przetestujemy sieć jako użytkownicy, aby zidentyfikować wybrane obrazy artystów, które nie są reprezentowane w zbiorze danych.

Ze względu na fakt, że zdjęcia mają różne rozmiary i rozdzielczości (w zestawie danych zdjęcia mają około 500 na 500 pikseli), zmniejszyliśmy wszystkie zdjęcia do jednego rozmiaru - 224 na 224 piksele. Także losowo obracamy i losowo przybliżamy obrazy dla naszej sieci. Ta losowość różnicuje dane treningowe i pomaga uniknąć przeuczenia, które może wystąpić ze względu na niewielki rozmiar zestawu danych.

Prosta CNN

Naszą pierwszą próbą rozwiązania była „klasyczna” konwolucyjna sieć neuronowa składająca się z:



Warto zauważyć, że kolor może być również ważnym elementem sztuki razem z formą, więc konwolucyjne warstwy używają filtrów dla każdej z map trzech głównych kolorów RGB.

Dla warstwy gęstej wykorzystaliśmy funkcji aktywacji „relu”, a dla warstwy wejścia - „softmax”

Dla komplikacji modelu sieci wykorzystano optimalizator Adam

Dla dopasywania modelu: funkcja kosztu – „Categorical Crossentropy”, a jako kryterium oceny- ‘accuracy’

Trzy najlepsze próby na zestawie testowym:

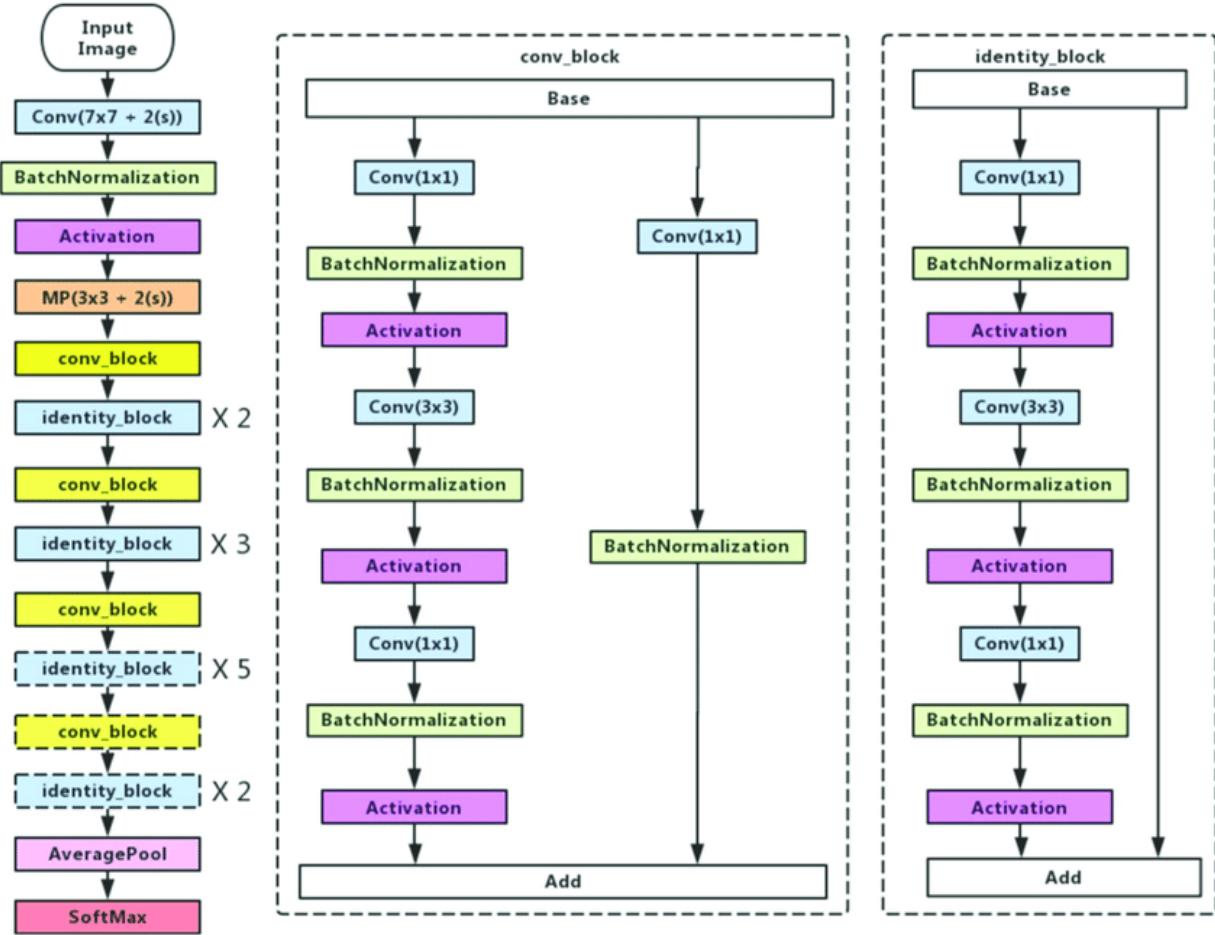
Nr epoch	val_loss	val_accuracy
10	7.04	0.34
8	7.26	0.34
4	7.45	0.3

Wyniki wydają się już całkiem dobre - opracowany algorytm znajdowania wspólnych cech na zdjęciach i klasyfikacji przedstawionych obiektów, jednoznacznie znajduje te same cechy na różnych obrazach artystów (średnia dokładność jest trzykrotnie wyższa niż dokładność dla losowego rozkładu odpowiedzi dla 10 klas – 0,1).

Ale cel wyznaczony na początku nie został jeszcze osiągnięty. Naszym największym problemem jest mały rozmiar zestawu danych i złożoność zadania. Potrzebna jest większa ilość warstw i większy zestaw danych. Dodanie dużej ilości nowych warstw zwiększy obciążenie komputera, a artyści nie zdążą już napisać nowych obrazów. W takim przypadku optymalnym rozwiązaniem byłoby użycie sieci, która została już przeszkolona w zakresie innego podobnego problemu. Takim rozwiązaniem została sieć **ResNet50** i zestaw wag **ImageNet**

ResNet50

ResNet-50 to konwolucyjna sieć neuronowa o głębokości 50 warstw. Wspiera możliwość załadowania wstępnie przeszkoloną wersję sieci na ponad milionie obrazów z bazy danych ImageNet. Wstępnie przeszkolona sieć może klasyfikować obrazy do 1000 kategorii obiektów, takich jak klawiatura, mysz, ołówek i wiele zwierząt.



Wstępne przetwarzanie danych dla tego modelu wygląda tak samo jak w poprzednim.

Wyjście z sieci musi być skierowane do warstwy GlobalAveragePooling2D, po czym dodajemy jedną w pełni połączoną warstwę 128 neuronów z funkcją aktywacji „relu” i ostatnią warstwę - 10 neuronów wyjściowych z aktywacją „softmax”

Kompilacja i dopasowywanie modelu podobne z poprzednimi – optimalizator Adam, funkcja kosztu – „Categorical Crossentropy”, kryterium oceny – ‘accuracy’.

Zobaczmy trzy najlepsze próby na zestawie testowym:

Nr epoch	val_loss	val_accuracy
3	1.02	0.63
2	1.11	0.64
5	1.29	0.64

Widzimy, że osiągnęliśmy i przekroczyliśmy nasz cel - 50%. Udało nam się wyszkolić sieć w zakresie identyfikacji i przewidywania autorów dzieł sztuki. Jeśli byśmy mieli

większą moc obliczeniową i dużego zbioru danych od różnych autorów można by było oczekwać nieco lepszego wyniku.

Na koniec dodajemy kilka przykładów tego, jak sieć przewiduje autorstwo, tak jakbyśmy byli użytkownikami korzystającymi z tej sieci w praktyce (na zielono podświetlona jest poprawna odpowiedź):



Simple CNN

- 0% Chagall
- 0% Dali
- 0% Gauguin
- 0% Hokusai
- 0% Malevich
- 100% **Picasso**
- 0% Raphael
- 0% Repin
- 0% Seurat
- 0% Warhol

0% **Chagall**
0% Dali
0% Gauguin
0% Hokusai
10% Malevich
0% Picasso
0% Raphael
0% Repin
0% Seurat
89% **Warhol**

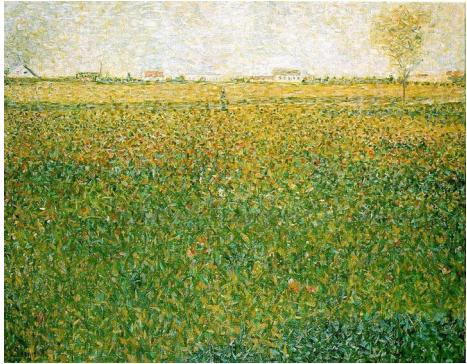
ResNet50

- 2% Chagall
- 0% Dali
- 0% Gauguin
- 0% Hokusai
- 0% Malevich
- 0% Picasso
- 0% Raphael
- 0% Repin
- 0% Seurat
- 97% **Warhol**

98% **Chagall**
0% Dali
1% Gauguin
0% Hokusai
0% Malevich
0% Picasso
0% Raphael
1% Repin
0% Seurat
0% Warhol

86% **Chagall**
0% Dali
0% Gauguin
14% Hokusai
0% Malevich
0% **Picasso**
0% Raphael
0% Repin
0% Seurat
0% Warhol

2% Chagall
0% Dali
1% Gauguin
0% Hokusai
0% Malevich
87% **Picasso**
0% Raphael
0% Repin
0% Seurat
10% Warhol



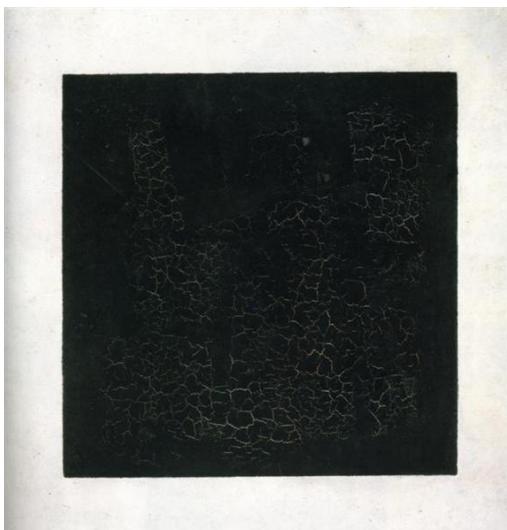
1% Chagall
0% Dali
0% Gauguin
0% Hokusai
0% Malevich
0% Picasso
0% Raphael
0% Repin
99% Seurat
0% Warhol

0% Chagall
0% Dali
0% Gauguin
0% Hokusai
0% Malevich
0% Picasso
0% Raphael
0% Repin
100% Seurat
0% Warhol



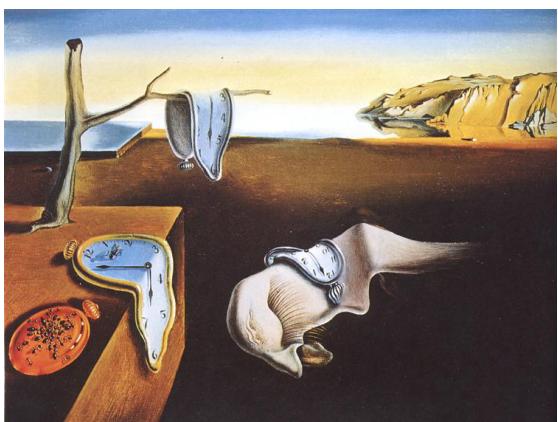
0% Chagall
0% Dali
0% Gauguin
0% Hokusai
0% Malevich
0% Picasso
0% Raphael
100% Repin
0% Seurat
0% Warhol

0% Chagall
0% Dali
6% Gauguin
0% Hokusai
0% Malevich
20% Picasso
0% Raphael
73% Repin
0% Seurat
0% Warhol



0% Chagall
0% Dali
0% Gauguin
0% Hokusai
100% Malevich
0% Picasso
0% Raphael
0% Repin
0% Seurat
0% Warhol

0% Chagall
13% Dali
0% Gauguin
1% Hokusai
85% Malevich
0% Picasso
0% Raphael
1% Repin
0% Seurat
1% Warhol



3% Chagall
0% Dali
0% Gauguin
0% Hokusai
0% Malevich
91% Picasso
4% Raphael
0% Repin
0% Seurat
3% Warhol

0% Chagall
4% Dali
0% Gauguin
1% Hokusai
0% Malevich
92% Picasso
0% Raphael
0% Repin
2% Seurat
1% Warhol