

Student IT/EE Workshop 2024

One-shot learning from prototype SKU images

Aleksandra Kowalczyk

Opiekun: dr inż. Grzegorz Sarwas

**Politechnika
Warszawska**

Warszawa, 18 kwietnia 2024 r.



Agenda

1

1. Systemy rozpoznawania półek sklepowych
2. Metody prototypowania
3. Wykorzystana metoda
4. Zbiór danych
5. Wyniki
6. Podsumowanie
7. Możliwości dalszego rozwoju



Systemy służące do rozpoznawania półek sklepowych

2

- zróżnicowane
- szeroko stosowane



Fig. 1: Kamery w sklepach samoobsługowych.



Fig. 2: Roboty podłogowe wykorzystywane w rozwiązaniu SCANDIT.



Fig. 3: Wirtualne półki sklepowe w aplikacji mobilnej w rozwiązaniu Omniaz.

SKU

- kod alfanumeryczny
- jednostka magazynująca
- odnosi się do konkretnego przedmiotu
- oznacza specyficzne cechy produktu
- przeznaczona do wewnętrznego użytku
- odpowiedni dla dynamicznej branży FMCG (dobra szybko zbywalne)



EN-PIL-B202



EN-GRL-A101

category-color,size-brand,series

Fig. 4: Przykład stworzenia SKU.



Metody prototypowania

- reprezentacja danych za pomocą zapamiętanych przykładów
- porównanie obserwacji z zestawem referencyjnych prototypów
- pokrywają się z koncepcjami stosowanymi w psychologii poznawczej

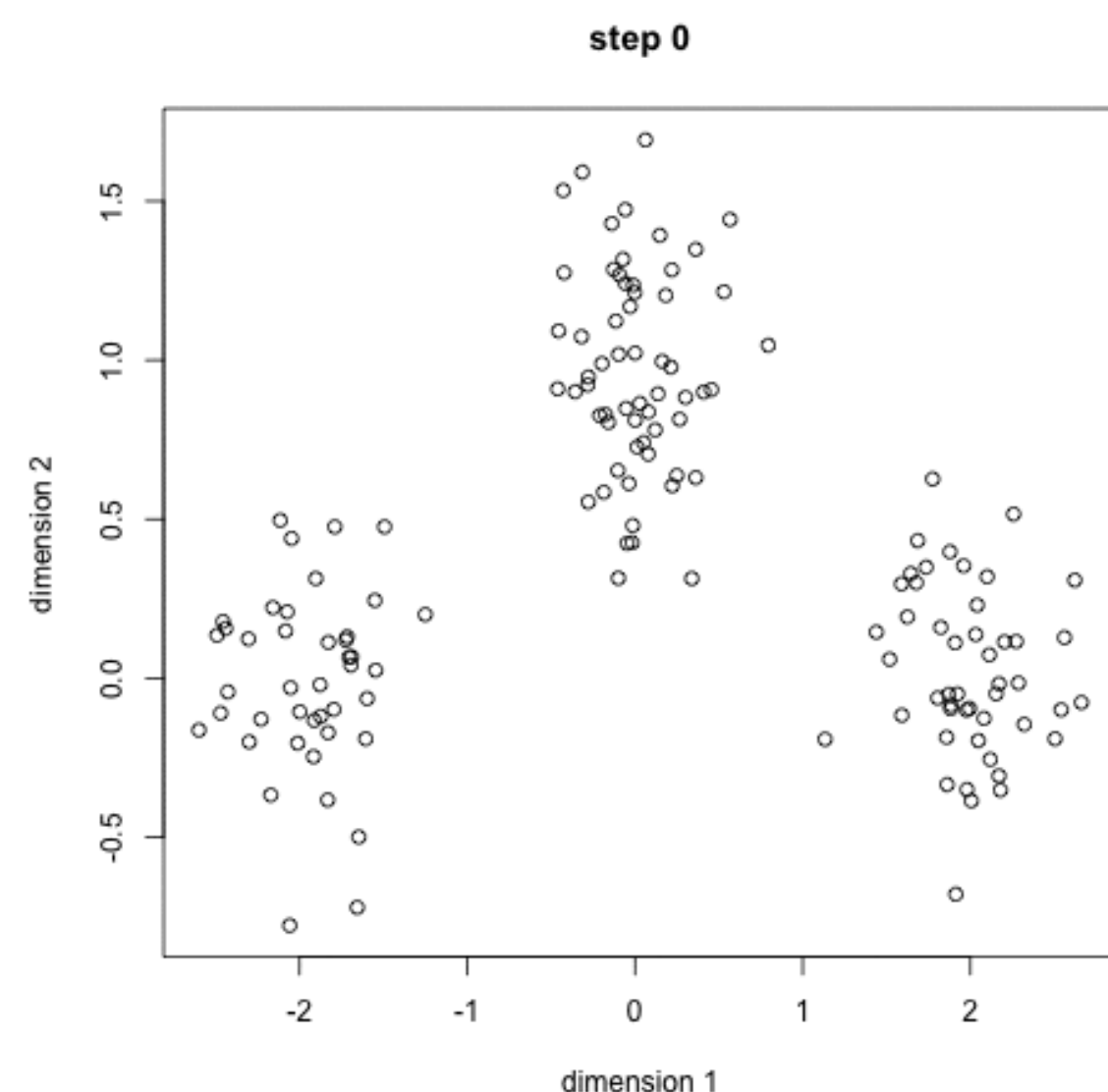


Fig. 5: Metoda k-means wyznaczająca centroid dla klastra obiektów jako średnią ważoną cech.

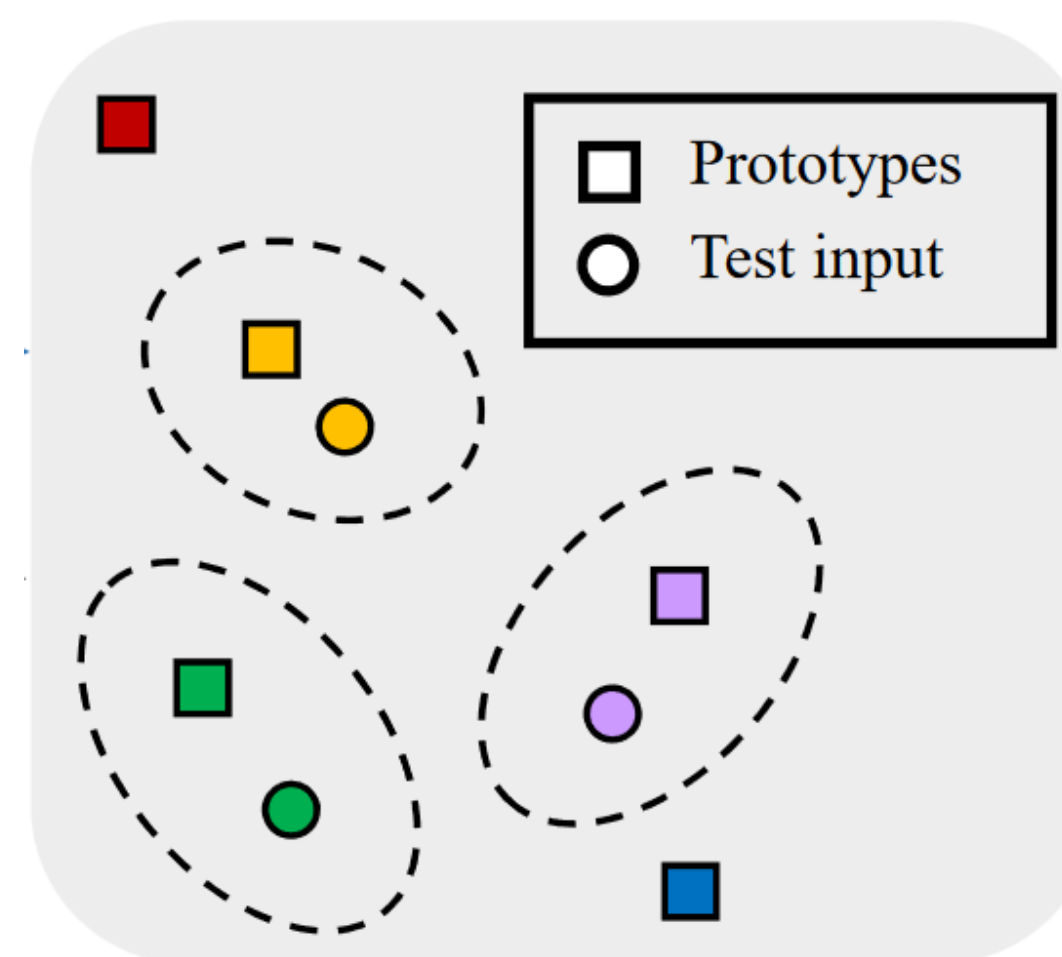


Fig. 6: Metoda w której prototyp jest na wstępie wyznaczony i niezmienny.

Inspiracją była praca „Variational Prototyping-Encoder: One-Shot Learning with Prototypical Images” (Kim i inni, 2019) przedstawiający wariacyjny autokoder do rozpoznawania znaków drogowych.

Wykorzystana metoda

5

Faza treningowa:

- koder przekształca obrazy i prototypy w rozkład ukryty
- dekoder odtwarza zakodowany rozkład w prototyp

Faza testowa:

- przeszkolony koder służy jako ekstraktor cech
- obrazy testowe i prototypy są kodowane do przestrzeni ukrytej
- klasyfikacja przy użyciu najbliższego sąsiada

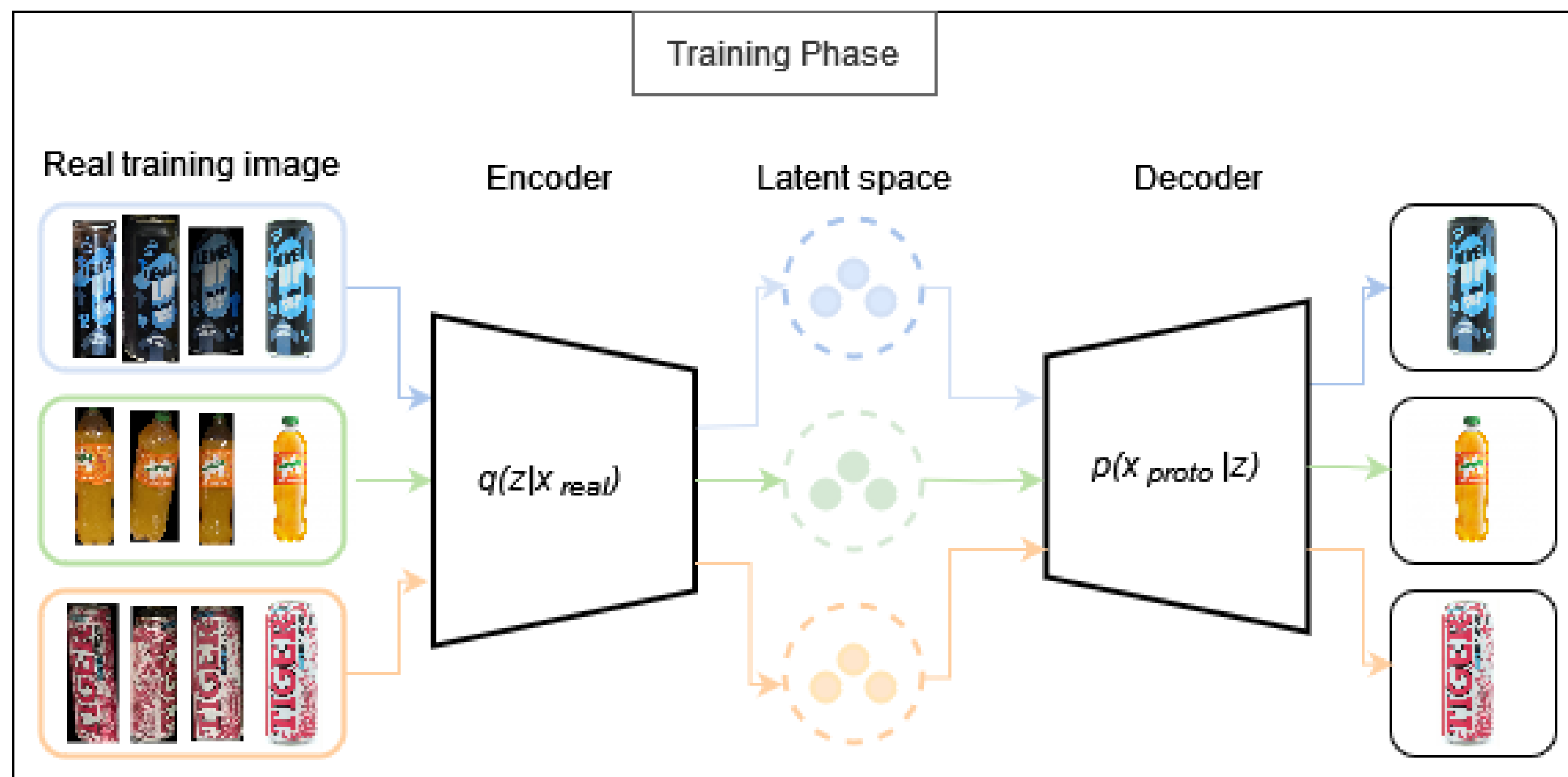


Fig. 7: Faza treningowa wariacyjnego autokodera.

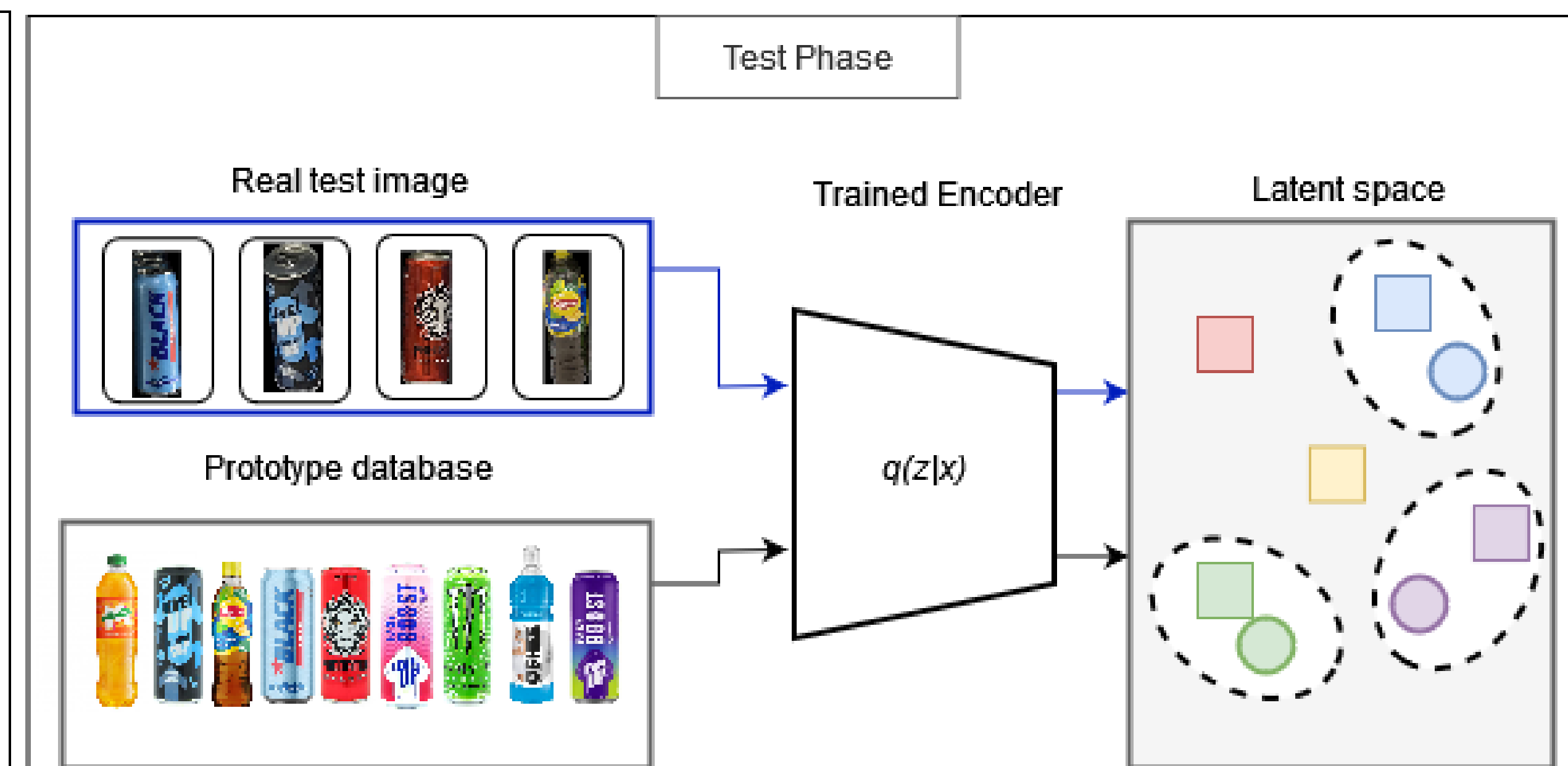


Fig. 8: Faza testowa wariacyjnego autokodera.

Definicja podobieństwa

6

Dystans Euklidesowy:

- pomiędzy dwoma wektorami \mathbf{u} i \mathbf{v} w przestrzeni \mathbf{R}^n
- $d(\mathbf{u}, \mathbf{v}) = \sqrt{\sum_{i=1}^n (\mathbf{u}_i - \mathbf{v}_i)^2}$
- Im niższy, tym większe podobieństwo pomiędzy wektorami

Podobieństwo Cosinusowe:

- mierzy kąt pomiędzy dwoma wektorami \mathbf{u} i \mathbf{v}
- $\cos(\mathbf{u}, \mathbf{v}) = \frac{\mathbf{u} \cdot \mathbf{v}}{\|\mathbf{u}\| \|\mathbf{v}\|}$, gdzie $\mathbf{u} \cdot \mathbf{v}$ oznacza iloczyn skalarny wektorów, a $\|\mathbf{u}\| \|\mathbf{v}\|$ są normami tych wektorów
- Im wyższy, tym większe podobieństwo pomiędzy wektorami



Zbiór danych

- 300 zdjęć półek sklepowych
- 10000 + wyodrębnionych produktów
- 50 klas napojów bezalkoholowych w puszkach i butelkach



Fig. 9: Przykładowe klasy z utworzonego zbioru danych.



Wyniki

- wpływ rozmiaru obrazu i technik augmentacji
- oddziaływanie transformera przestrzennego



Distance	Method	Recall			Top-nn	
		All	Train	Test	2-nn	3-nn
Euclidean	200 epochs	0.888	0.894	0.883	0.972	0.986
	Trigger by val	0.769	0.939	0.623	0.825	0.839
Cosine	200 epochs	0.916	0.909	0.922	0.986	0.993
	Trigger by val	0.888	0.955	0.831	0.986	0.986

Tab. 1: Porównanie osiągniętych wartości czułości dla badanych odległości po osiągnięciu zadeklarowanej ilości epok w procesie uczenia lub metodą wczesnego zatrzymania (ang. Early stopping).

Image size	Algorithm's version	One-shot classification recall (%)	
		Classes seen	Classes unseen
48 × 48	VPE	0.939	0.961
	VPE + aug	0.939	0.896
	VPE + aug + rotate	0.576	0.818
	VPE + stn	0.939	0.948
	VPE + aug + stn	0.955	0.896
64 × 64	VPE	0.924	0.740
	VPE + aug	0.970	0.909
	VPE + aug + rotate	0.712	0.909
	VPE + stn	0.939	0.935
	VPE + aug + stn	0.909	0.922

Tab. 2: Jednorazowa klasyfikacja dla różnych wersji algorytmu i rozmiaru obrazu.

Osiągnięte rezultaty

- Wyzwania
 - rozróżnienie produktów tej samej marki o różnych wariantach smakowych
 - rozróżnienie produktów różnych firm o tych samych kolorach

- Ocena jakości modelu:

- czułość $Recall = \frac{TP}{TP+FN}$
- dokładność $Accuracy = \frac{TP+TN}{P+N}$
- precyzja $Precision = \frac{TP}{TP+FP}$

Classes	Recall	Accuracy	Precision
Seen			
Black orange	1.000	1.000	1.000
Coca-cola	1.000	1.000	1.000
Easy boost pink	0.818	0.986	1.000
Easy boost purple	0.909	0.993	1.000
Level up blue	1.000	0.993	0.917
Dzik green	1.000	0.979	0.786
Unseen			
Black light-blue	0.636	0.972	1.000
Monster light-blue	1.000	0.972	0.733
Monster pink	1.000	0.986	0.846
Black green	0.727	0.979	1.000
Predator red	1.000	1.000	1.000
Lipton bottle lemon	1.000	1.000	1.000
Oshee bottle blue	1.000	1.000	1.000

Tab. 3: Osiągnięte wartości czułości, dokładności i precyzji dla klas testowych.

Wizualizacja rezultatów

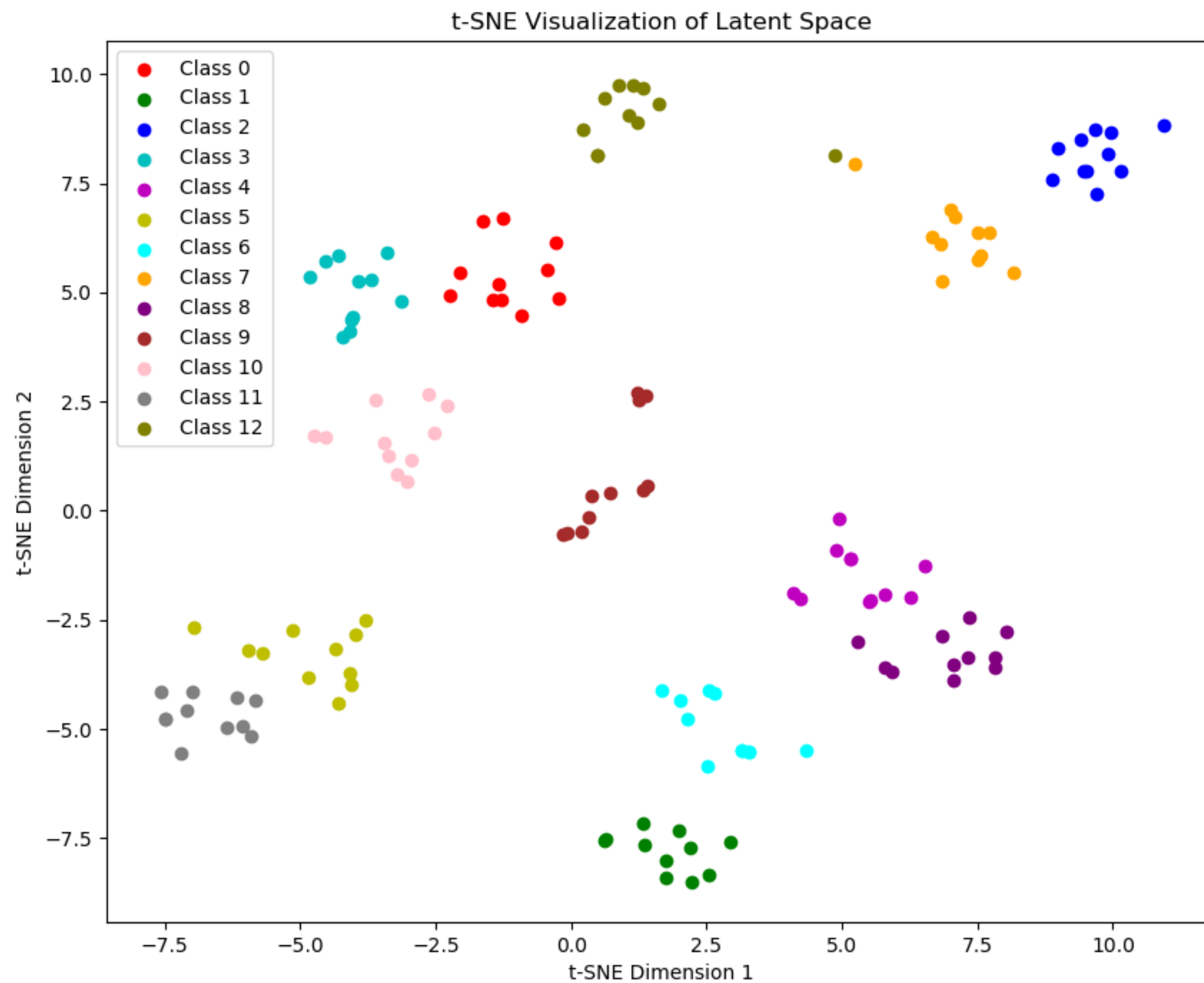


Fig. 10: Wizualizacja t-SNE dla zbioru testowego po wykonaniu 150 epok na zbiorze uczącym.



Fig. 11: Prototypy klas wraz z rekonstrukcjami wykonanymi przez koder wykonane po 150 epokach uczących, gdy uzyskana wydajność przekracza 90% zarówno dla klas produktów znanych przez model, jak i nieznanymi.

Podsumowanie

11

- Wariacyjny autokoder został zastosowany do rozpoznawania napojów na półkach sklepowych
- Metoda wykorzystująca jedną grafikę internetową produktu jako prototyp okazała się skuteczna
- Przetestowany model prawidłowo rozróżnia produkty o tym samym kolorze oraz różne smaki tego samego napoju
- Model zapewnia pewien poziom odporności na różnice wizualne pomiędzy produktami oraz ich wariantami w ramach tych samych kategorii



Możliwości dalszego rozwoju

12

- dostosowanie i przetestowanie modelu na produktach innych kategorii
- dalsze prace i testy przy zaimplementowanej sieci neuronowej w celu osiągnięcia większej dokładności dla branży spożywczej



Dziękuję za uwagę.

