

# Metody kontrastowe w problemach abstrakcyjnego wnioskowania

Contrastive methods in abstract visual reasoning

---

Mikołaj Małkiński

20 stycznia 2021

Politechnika Warszawska

Wydział Matematyki i Nauk Informatycznych

1. Motywacja
2. Tworzenie analogii poprzez kontrastowanie (LABC)
3. Budowa reprezentacji obrazów poprzez kontrastowanie
4. Kontrastowanie w nauce rozwiązywania RPM
5. Podsumowanie

# Motywacja

---

Zadanie 1. Oblicz  $\sqrt{27}$

a).  $-5$

b).  $0$

c).  $3\sqrt{3}$

d).  $20\frac{1}{4}$

Zadanie 1. Oblicz  $\sqrt{27}$

a).  $2\sqrt{3}$

b).  $3\frac{1}{3}$

c).  $3\sqrt{3}$

d). 9

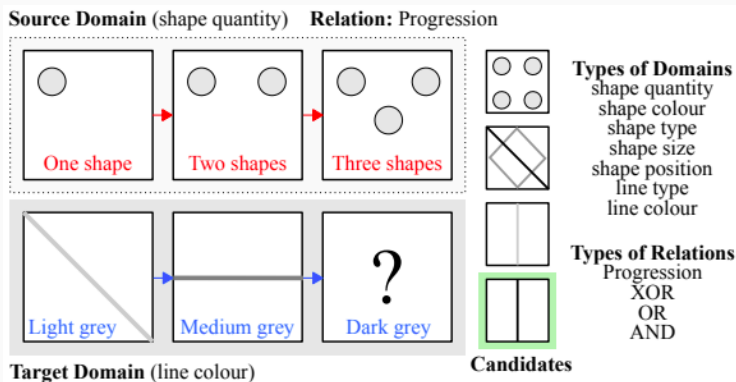
# Tworzenie analogii poprzez kontrastowanie (LABC)

---

## Tworzenie analogii poprzez kontrastowanie (LABC)

Tworzenie analogii wymaga reprezentacji relacyjnych struktur w sposób umożliwiający zastosowanie ich do różnych domen Hill et al. [2019]

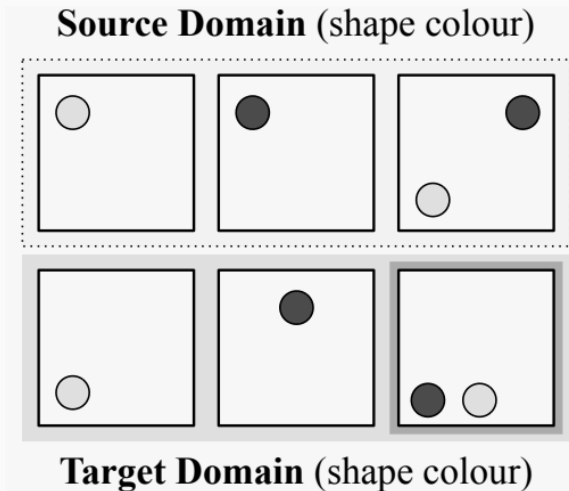
# Tworzenie analogii poprzez kontrastowanie (LABC)



Rysunek 1: Problem tworzenia analogii wizualnych z [Hill et al., 2019].

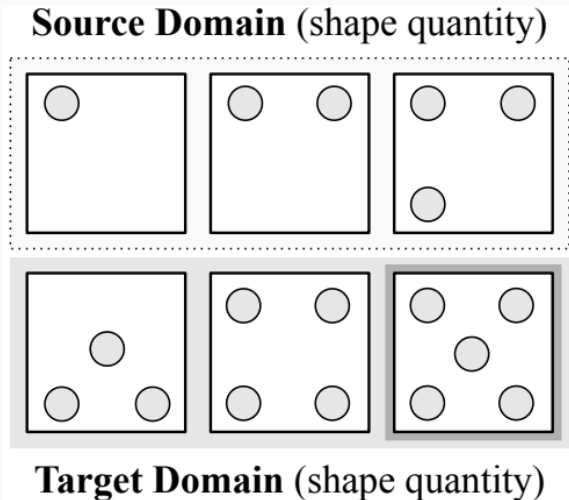


## Tworzenie analogii poprzez kontrastowanie (LABC)



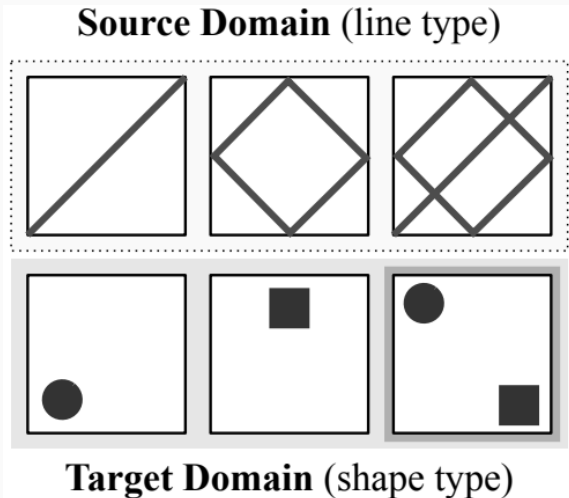
**Rysunek 2:** Matryca z tą samą domeną źródłową i docelową oraz tymi samymi wartościami.

# Tworzenie analogii poprzez kontrastowanie (LABC)



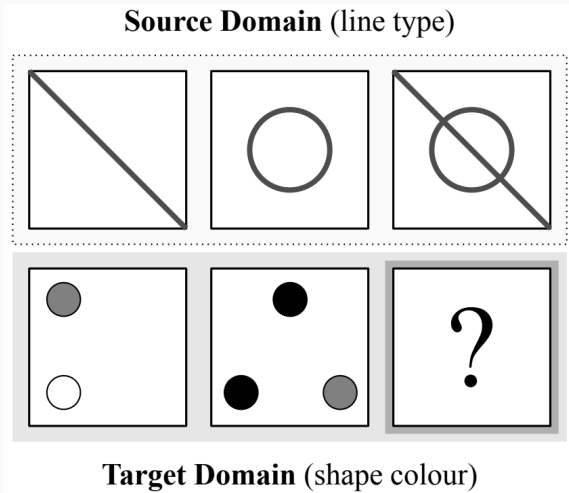
**Rysunek 3:** Matryca z tą samą domeną źródłową i docelową oraz innymi wartościami.

## Tworzenie analogii poprzez kontrastowanie (LABC)



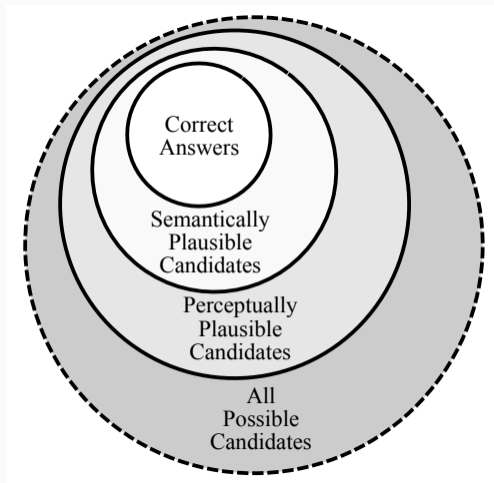
**Rysunek 4:** Matryca z inną domeną źródłową i docelową oraz innymi wartościami.

# Tworzenie analogii poprzez kontrastowanie (LABC)



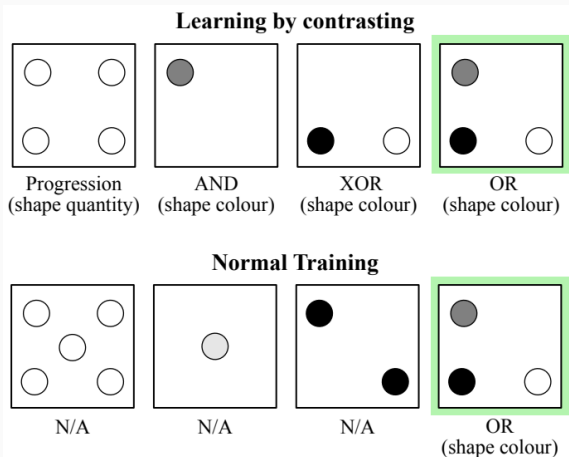
**Rysunek 5:** Matryca z inną domeną źródłową i docelową oraz innymi wartościami z brakującą odpowiedzią.

## Tworzenie analogii poprzez kontrastowanie (LABC)



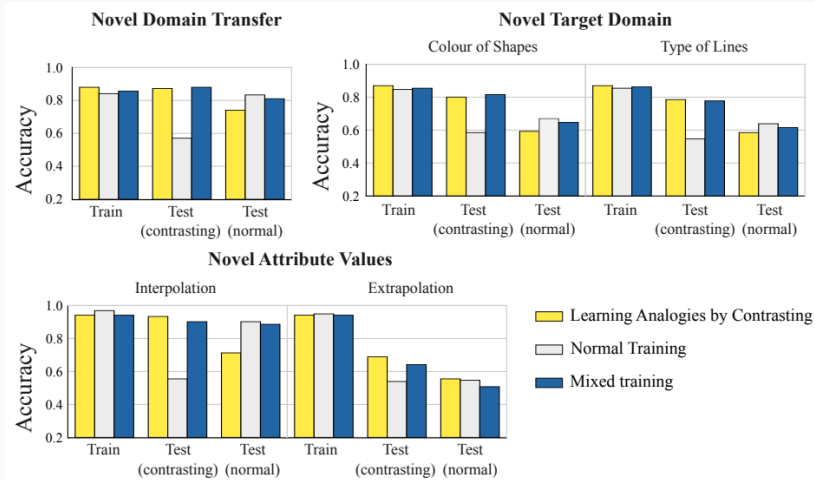
**Rysunek 6:** Grupy możliwych odpowiedzi w zadaniach abstrakcyjnego wnioskowania z [Hill et al., 2019].

# Tworzenie analogii poprzez kontrastowanie (LABC)



**Rysunek 7:** Metoda generowania możliwych odpowiedzi poprzez kontrastowanie. Błędne odpowiedzi są nie tylko wizualnie, ale też **semantycznie prawdopodobne**.

# Tworzenie analogii poprzez kontrastowanie (LABC)



Rysunek 8: Wyniki metody LABC.

## Tworzenie analogii poprzez kontrastowanie (LABC)

- LABC zwiększa zdolności generalizacji ponad treningowy rozkład danych.
- Efekt zachodzi w procesach tworzenia analogii pomiędzy nowymi domenami.
- Metoda daje umiarkowaną poprawę zdolności ekstrapolacji poza wartości ze zbioru treningowego.



## Tworzenie analogii poprzez kontrastowanie (LABC)

LABC wymaga znajomości zachodzących abstrakcyjnych struktur.

Alternatywne metody:

- topk: 1) wygenerowanie wielu możliwych odpowiedzi, 2) wybranie top- $k$  najwyżżej ocenionych odpowiedzi, 3) wykonanie wstecznej propagacji błędu tylko dla wybranych odpowiedzi,
- adversarial: użycie modelu generatora do tworzenia trudnych przykładów.

## Tworzenie analogii poprzez kontrastowanie (LABC)

Training method	Test accuracy
LABC; explicit SMT	0.89
LABC; top-k	0.77
LABC; adversarial	0.62
Random candidate answers	0.25

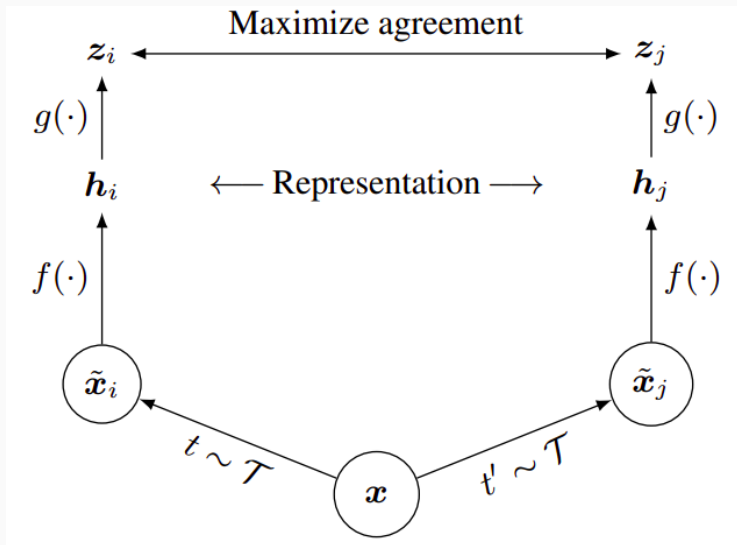
Rysunek 9: Wyniki metody LABC i alternatyw.

Metody adwersarialne (GAN Goodfellow et al. [2014]) i samodzielnej gry (self-play Silver et al. [2017]) mogą być uważane za automatyczną aproksymację LABC.

# Budowa reprezentacji obrazów poprzez kontrastowanie

---

W jaki sposób zbudować model do reprezentacji obrazów nie mając dostępu do etykiet?



Rysunek 10: Metoda SimCLR Chen et al. [2020].

---

**Algorithm 1** SimCLR's main learning algorithm.

---

**input:** batch size  $N$ , constant  $\tau$ , structure of  $f, g, \mathcal{T}$ .  
**for** sampled minibatch  $\{\mathbf{x}_k\}_{k=1}^N$  **do**  
  **for all**  $k \in \{1, \dots, N\}$  **do**  
    draw two augmentation functions  $t \sim \mathcal{T}, t' \sim \mathcal{T}$   
    # the first augmentation  
     $\tilde{\mathbf{x}}_{2k-1} = t(\mathbf{x}_k)$   
     $\mathbf{h}_{2k-1} = f(\tilde{\mathbf{x}}_{2k-1})$  # representation  
     $\mathbf{z}_{2k-1} = g(\mathbf{h}_{2k-1})$  # projection  
    # the second augmentation  
     $\tilde{\mathbf{x}}_{2k} = t'(\mathbf{x}_k)$   
     $\mathbf{h}_{2k} = f(\tilde{\mathbf{x}}_{2k})$  # representation  
     $\mathbf{z}_{2k} = g(\mathbf{h}_{2k})$  # projection  
  **end for**  
  **for all**  $i \in \{1, \dots, 2N\}$  and  $j \in \{1, \dots, 2N\}$  **do**  
     $s_{i,j} = \mathbf{z}_i^\top \mathbf{z}_j / (\|\mathbf{z}_i\| \|\mathbf{z}_j\|)$  # pairwise similarity  
  **end for**  
  **define**  $\ell(i, j)$  **as**  $\ell(i, j) = -\log \frac{\exp(s_{i,j}/\tau)}{\sum_{k=1}^{2N} \mathbb{1}_{[k \neq i]} \exp(s_{i,k}/\tau)}$   
   $\mathcal{L} = \frac{1}{2N} \sum_{k=1}^N [\ell(2k-1, 2k) + \ell(2k, 2k-1)]$   
  update networks  $f$  and  $g$  to minimize  $\mathcal{L}$   
**end for**  
**return** encoder network  $f(\cdot)$ , and throw away  $g(\cdot)$

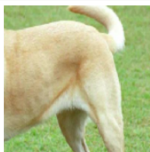
---

**Rysunek 11:** Algorytm SimCLR.





(a) Original



(b) Crop and resize



(c) Crop, resize (and flip)



(d) Color distort. (drop)



(e) Color distort. (jitter)



(f) Rotate  $\{90^\circ, 180^\circ, 270^\circ\}$



(g) Cutout



(h) Gaussian noise



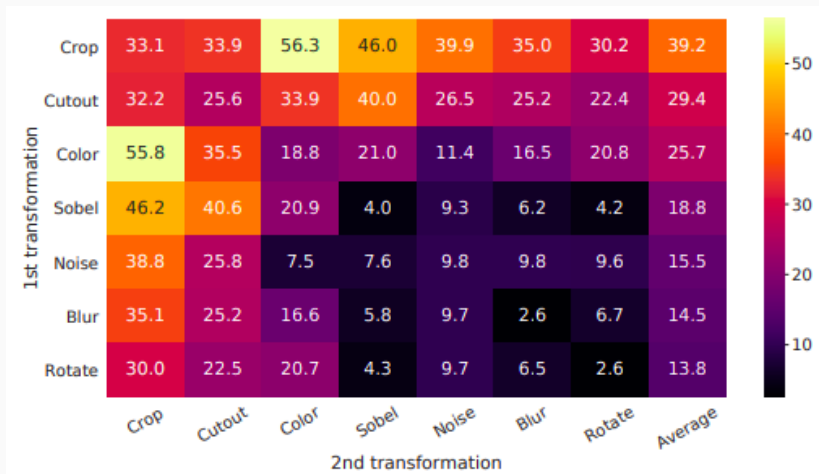
(i) Gaussian blur



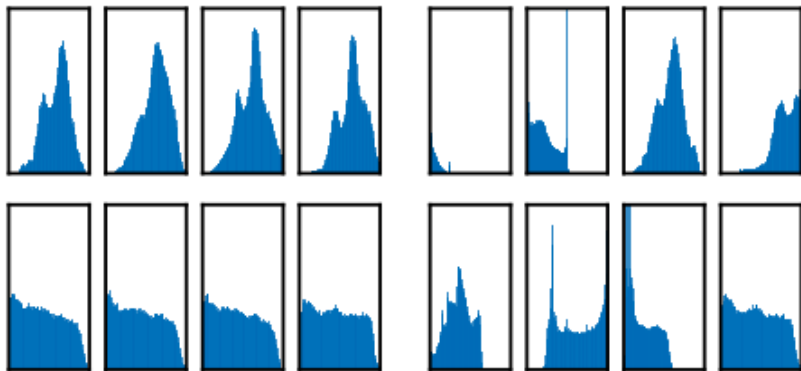
(j) Sobel filtering

**Rysunek 12:** Augmentacje obrazów.

# SimCLR



**Rysunek 13:** Wyniki augmentacji obrazów (ImageNet top-1 używając liniowej ewaluacji).

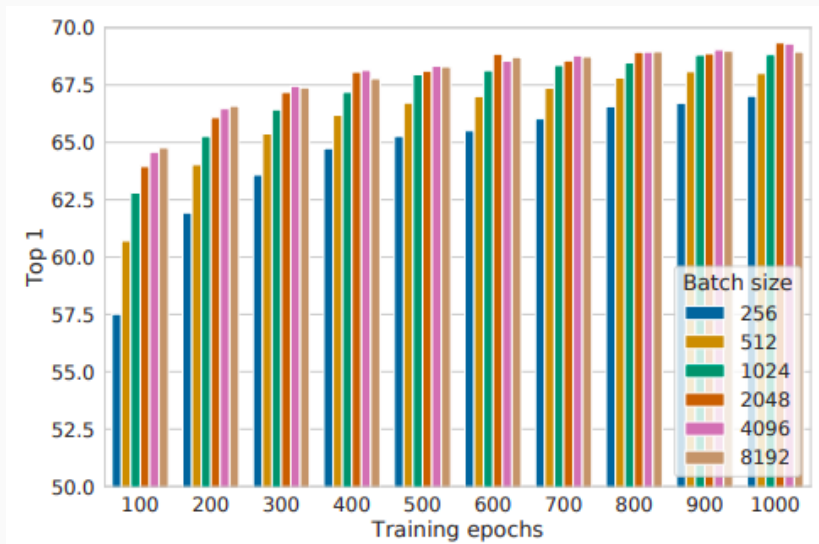


(a) Without color distortion.

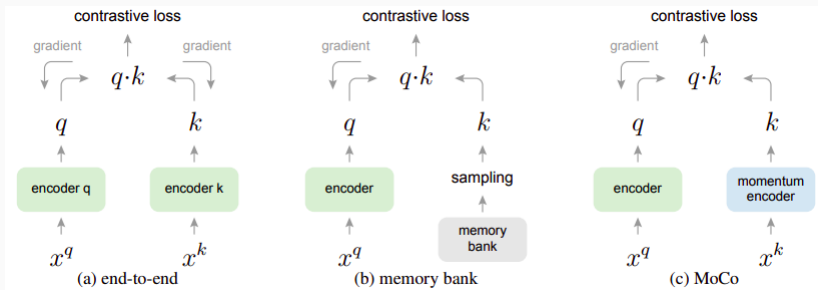
(b) With color distortion.

**Rysunek 14:** Histogramy intensywności pixeli używając wycinków dwóch różnych obrazów.

# SimCLR



Rysunek 15: SimCLR wymaga dużego rozmiaru batcha.

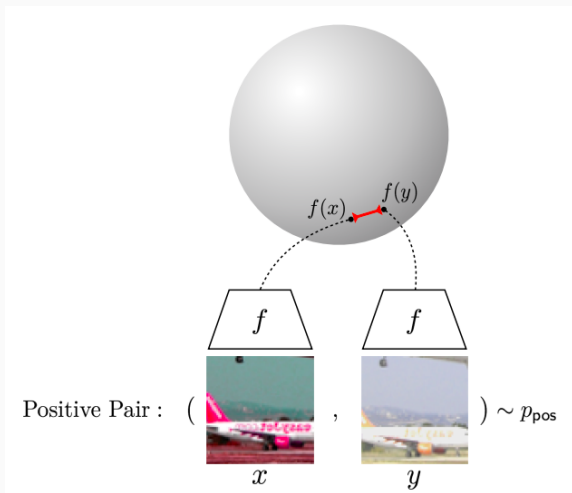


**Rysunek 16:** Mechanizmy kontrastowania zwiększające liczbę dostępnych próbek He et al. [2020].

Jakie reprezentacje otrzymujemy wykorzystując naukę poprzez kontrastowanie?

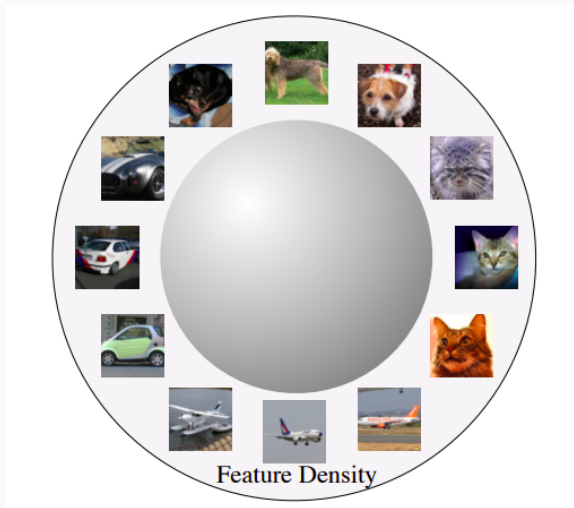
Nauka poprzez kontrastowanie asymptotycznie optymalizuje wyrównanie (alignment) i jednolitość (uniformity) cech na hipersferze Wang and Isola [2020].

# Reprezentacje na hipersferze

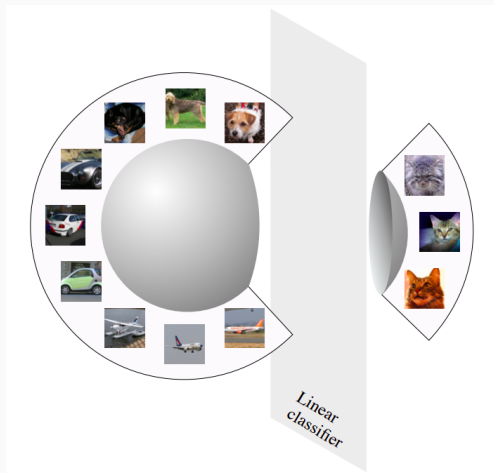


**Rysunek 17:** Podobne obserwacje mają podobne reprezentacje – wyrównanie na hipersferze (alignment on the hypersphere) Wang and Isola [2020].





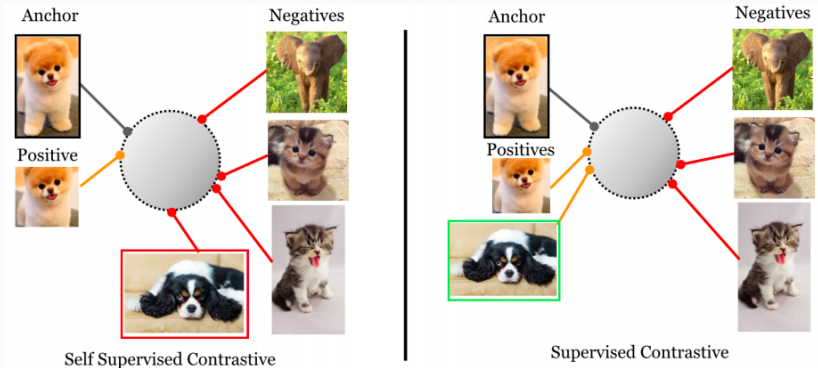
**Rysunek 18:** Klasy są równomiernie rozłożone – jednolitość na hipersferze (uniformity on the hypersphere) Wang and Isola [2020].



**Rysunek 19:** Dobrze pogrupowane klastry są liniowo separowalne Wang and Isola [2020].

Co jeśli mamy dostęp do etykiet?

# Nadzorowane uczenie poprzez kontrastowanie



**Rysunek 20:** Nienadzorowana vs nadzorowana nauka poprzez kontrastowanie Khosla et al. [2020].

$$\mathcal{L}^{\text{sup}} = \sum_{i=1}^{2N} \mathcal{L}_i^{\text{sup}} \quad (1)$$

$$\mathcal{L}^{\text{sup}} = \sum_{i=1}^{2N} \mathcal{L}_i^{\text{sup}} \quad (1)$$

$$\mathcal{L}_i^{\text{sup}} = \frac{1}{2N_{\tilde{Y}_i} - 1} \sum_{j=1}^{2N} \mathbb{1}_{i \neq j} \cdot \mathbb{1}_{\tilde{Y}_i = \tilde{Y}_j} \cdot \mathcal{L}_{i,j}^{\text{sup}} \quad (2)$$

$$\mathcal{L}^{\text{sup}} = \sum_{i=1}^{2N} \mathcal{L}_i^{\text{sup}} \quad (1)$$

$$\mathcal{L}_i^{\text{sup}} = \frac{1}{2N_{\tilde{Y}_i} - 1} \sum_{j=1}^{2N} \mathbb{1}_{i \neq j} \cdot \mathbb{1}_{\tilde{Y}_i = \tilde{Y}_j} \cdot \mathcal{L}_{i,j}^{\text{sup}} \quad (2)$$

$$\mathcal{L}_{i,j}^{\text{sup}} = -\log \frac{\exp(z_i \cdot z_j / \tau)}{\sum_{k=1}^{2N} \mathbb{1}_{i \neq k} \cdot \exp(z_i \cdot z_k / \tau)} \quad (3)$$

# Nadzorowane uczenie poprzez kontrastowanie

Loss	Architecture	Augmentation	Top-1	Top-5
Cross-Entropy (baseline)	ResNet-50	MixUp [61]	77.4	93.6
Cross-Entropy (baseline)	ResNet-50	CutMix [60]	78.6	94.1
Cross-Entropy (baseline)	ResNet-50	AutoAugment [5]	78.2	92.9
Cross-Entropy (our impl.)	ResNet-50	AutoAugment [30]	77.6	95.3
SupCon	ResNet-50	AutoAugment [5]	<b>78.7</b>	<b>94.3</b>
Cross-Entropy (baseline)	ResNet-200	AutoAugment [5]	80.6	95.3
Cross-Entropy (our impl.)	ResNet-200	Stacked RandAugment [49]	80.9	95.2
SupCon	ResNet-200	Stacked RandAugment [49]	<b>81.4</b>	<b>95.9</b>
SupCon	ResNet-101	Stacked RandAugment [49]	80.2	94.7

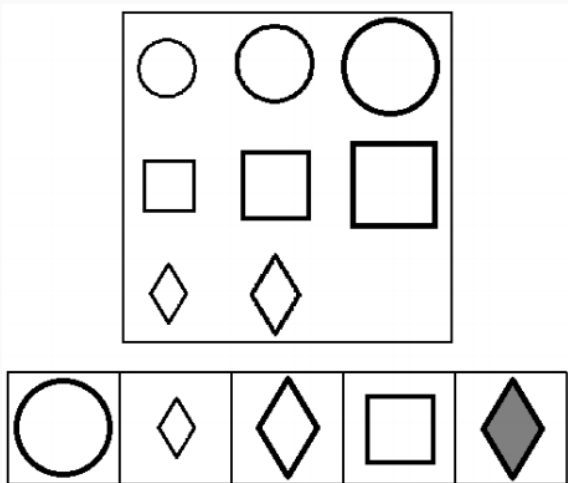
**Rysunek 21:** Wyniki na zbiorze ImageNet Khosla et al. [2020].



# Kontrastowanie w nauce rozwiązywania RPM

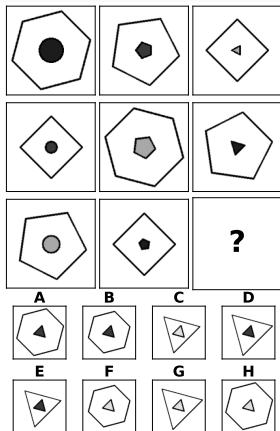
---

## Progresywne Matryce Ravena



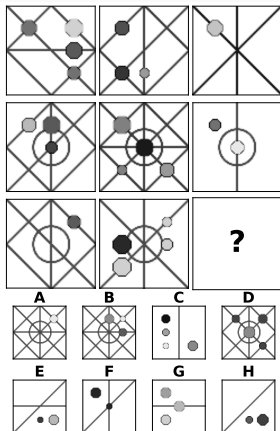
**Rysunek 22:** Przykład progresywnej matrycy Ravena (RPM) z [Mańdziuk and Żychowski, 2019] - dobrą odpowiedzią jest trzecia figura.

# Zbiór danych I-RAVEN



**Rysunek 23:** Przykład RPM ze zbioru danych I-RAVEN [Zhang et al., 2019], którą opisuje aż 8 różnych relacji dotyczących wewnętrznej i zewnętrznej struktury. Poprawna odpowiedź to A.

# Zbiór danych PGM



**Rysunek 24:** Przykład RPM ze zbioru danych PGM [Barrett et al., 2018], w której zachodzi tylko jedna reguła—operacja AND zastosowana do pozycji kształtów w wierszach. Panel A jest poprawną odpowiedzią.

# Definicja problemu

Niech  $\mathcal{M} = \{X, Y, k\}$  będzie progresywną matrycą Ravena:

- $X = \{x_i\}_{i=1\dots 16}$  - zbiór paneli/obrazów,
- $Y \subset \mathcal{Y}$  - zbiór reguł taki, że  $1 \leq |Y| \leq N_y$ ,
- $k \in \{1 \dots 8\}$  - indeks prawidłowej odpowiedzi.

W każdej RPM możemy zidentyfikować:

- kontekst matrycy  $\mathcal{X} = \{x_i\}_{i=1\dots 8} \subset X$ ,
- zbiór odpowiedzi  $\mathcal{C} = \{x_i\}_{i=9\dots 16} \subset X$ .

# Generalna definicja modelu rozwiązującego RPM

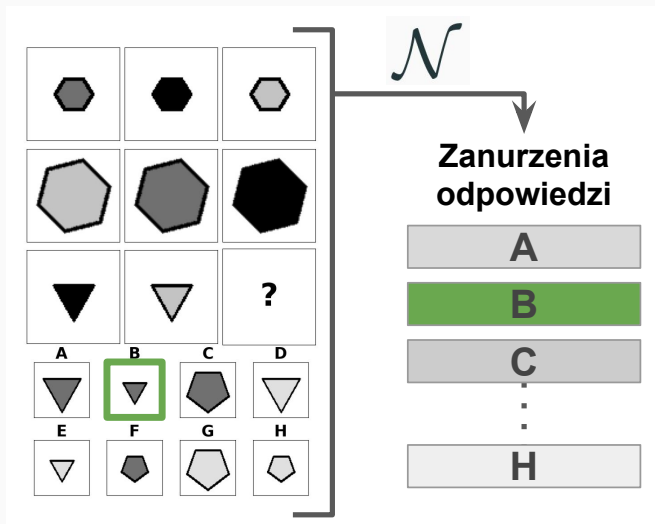
Można zidentyfikować dwa główne komponenty modelu do rozwiązywania RPM:

- komponent percepcyjny  $\mathcal{N}_v(x) = \mathcal{P} \in \mathbb{R}^{n \times d_p}$ ,
- komponent wnioskujący  $\mathcal{N}_r(\{\mathcal{P}_i\}_{i \in \{1, \dots, 8\} \cup \{k\}}) = h_k \in \mathbb{R}^{d_h}$ .

Są one zintegrowane w model  $\mathcal{N} = \mathcal{N}_v \circ \mathcal{N}_r$ , zdefiniowany jako  $\mathcal{N}(X) = \{h_k\}_{k=1..8}$ .

Wektor  $h$  został nazwany *zanurzeniem odpowiedzi* (embedding / ukryta reprezentacja).

# Ilustracja modelu rozwiązującego RPM



Rysunek 25: Ilustracja modelu rozwiązującego RPM, który generuje zanurzenia odpowiedzi.

## Trening nadzorowany (CE)

Prawdziwy rozkład prawdopodobieństwa  $p = \text{onehot}(k)$  jest estymowany poprzez:

- moduł  $\psi(h) = s \in \mathbb{R}$  ocenia każdą odpowiedź,
- oceny są zamieniane w rozkład prawdopodobieństwa  $\hat{p} = \text{softmax}(\{s_k\}_{k=1\dots 8})$ .

Cały model  $\mathcal{N} \circ \psi$  jest optymalizowany względem funkcji straty cross entropy  $\mathcal{L}^{\text{ce}}$ .



## Trening pomocniczy (AUX)

Prawdziwe reguły  $Y$  są enkodowane jako  $\pi = \text{encode}(Y)$ . Model  $\rho$  przewiduje reguły na podstawie zanurzeń odpowiedzi  $\hat{\pi} = \rho(\{h_k\}_{k=1\dots 8}) \in \mathbb{R}^{d_\pi}$ . Reguły są kodowane dwoma metodami:

- kodowanie gęste [Barrett et al., 2018, Zhang et al., 2019],
- kodowanie rzadkie (autorski sposób).

Binarna funkcja cross entropy oblicza błąd modelu  $\mathcal{L}^{\text{aux}}$ , który stanowi część łączonej funkcji kosztu:

$$\mathcal{L} = \mathcal{L}^{\text{ce}} + \beta \mathcal{L}^{\text{aux}} \quad (4)$$

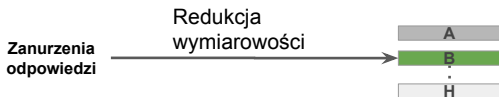
# Wyniki treningu nadzorowanego i pomocniczego

Metoda		Skuteczność (%)
		$\mu \pm \sigma$
SCL	CE	$82.8 \pm 0.7$
	AUX-dense	$86.9 \pm 1.5$
	AUX-sparse	<b><math>95.6 \pm 0.6</math></b>
SRAN	CE	$56.0 \pm 1.8$
	AUX-dense	$59.4 \pm 0.7$
	AUX-sparse	<b><math>59.9 \pm 1.6</math></b>
CoPINet	CE	$44.8 \pm 0.8$
	AUX-dense	$32.2 \pm 1.9$
	AUX-sparse	<b><math>50.3 \pm 1.0</math></b>

**Tablica 1:** Skuteczność testowa modeli opcjonalnie wspieranych treningiem pomocniczym na zbiorze danych I-RAVEN uśredniona z 4 przebiegów z losową inicjalizacją.

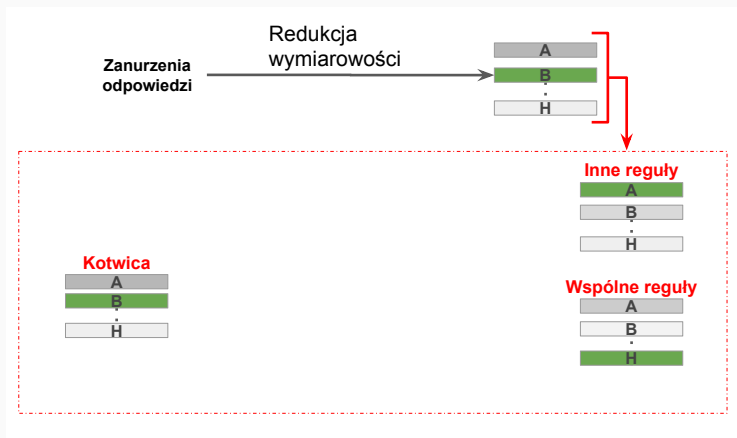
Jak zwiększyć nacisk na wykrywanie reguł?

# Trening kontrastujący (MLCL)



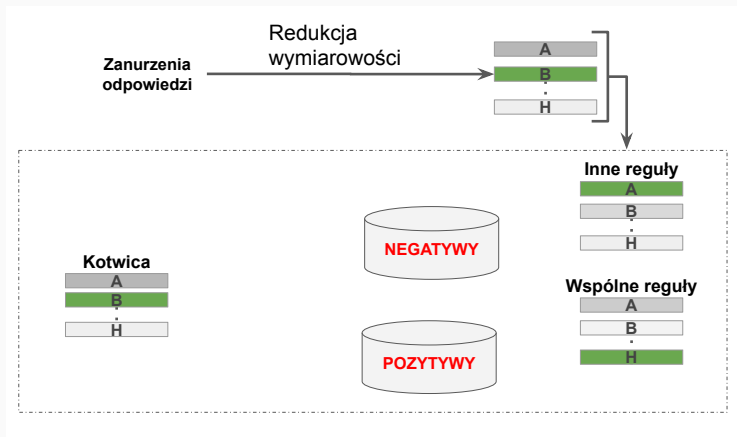
**Rysunek 26:** Moduł  $g$  zmniejsza wymiarowość zanurzeń odpowiedzi, generując *zanurzenia kontrastowe*  $z = g(h)$ . Wektory  $z$  są  $\ell_2$ -normalizowane, aby móc policzyć podobieństwo między  $z_i$  a  $z_j$  używając ich iloczynu skalarnego.

# Trening kontrastujący (MLCL)



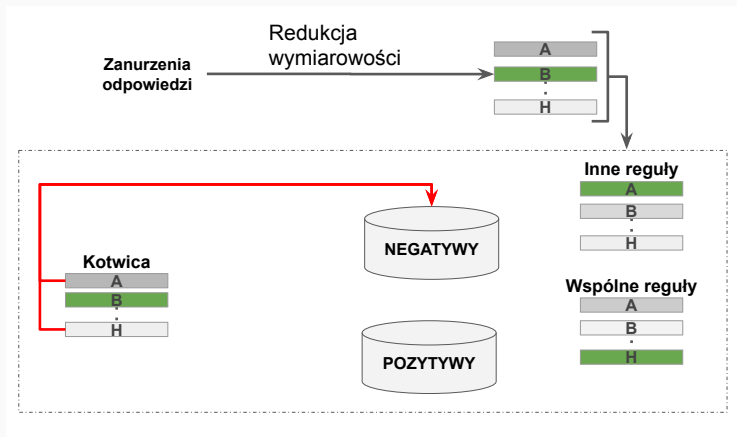
**Rysunek 27:** Dla każdej próbki (kotwicy) w danym batchu, oddzielnie rozważane są pozostałe obserwacje z innymi i ze wspólnymi regułami.

# Trening kontrastujący (MLCL)



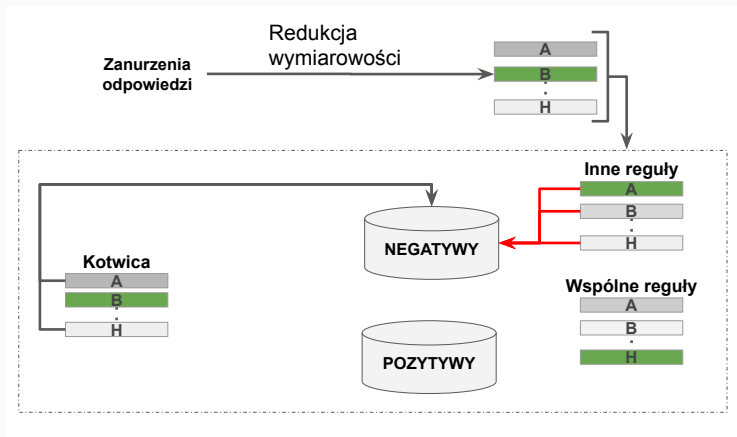
**Rysunek 28:** Obserwacje różne od kotwicy dzielone są na negatywne i pozytywne.

# Trening kontrastujący (MLCL)



**Rysunek 29:** Do zbioru negatywnych obserwacji zaliczane są błędnie uzupełnione matryce wygenerowane z kotwicy.

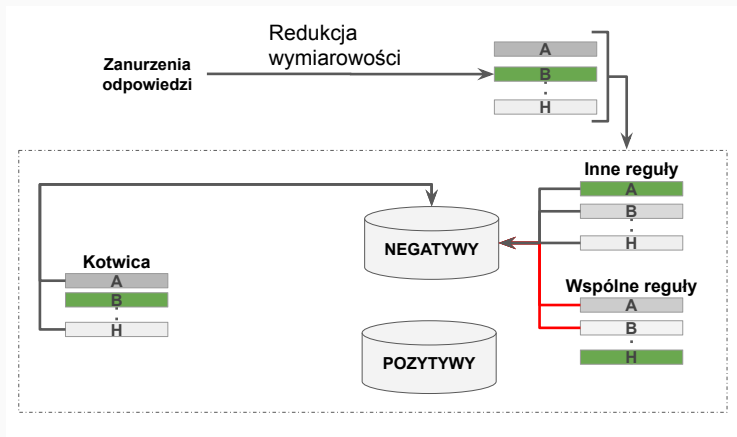
# Trening kontrastujący (MLCL)



**Rysunek 30:** Do zbioru negatywnych obserwacji zaliczane są wszystkie uzupełnienia matryc z innymi regułami.

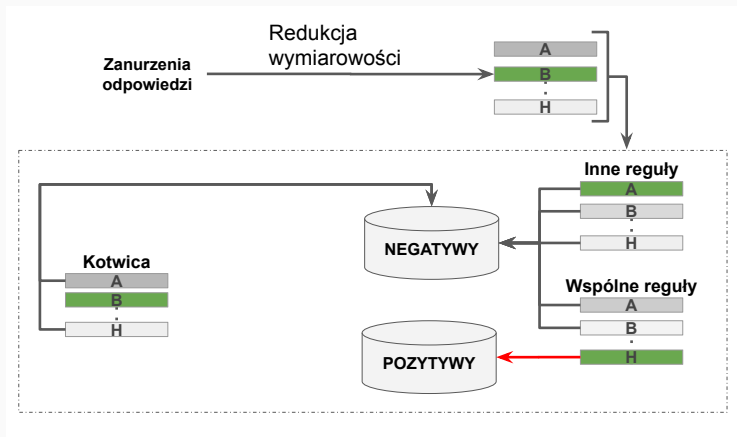


# Trening kontrastujący (MLCL)



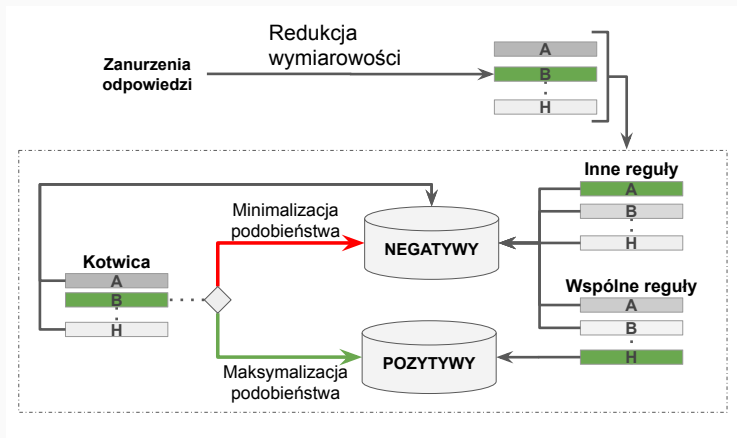
**Rysunek 31:** Do zbioru negatywnych obserwacji zaliczane są błędne uzupełnienia matryc ze wspólnymi regułami.

# Trening kontrastujący (MLCL)



**Rysunek 32:** Do zbioru pozytywnych obserwacji zaliczane są poprawne uzupełnienia matryc ze wspólnymi regułami.

# Trening kontrastujący (MLCL)



**Rysunek 33:** Nauka polega na minimalizacji podobieństwa kotwicy do przykładów negatywnych oraz maksymalizacji podobieństwa do przykładów pozytywnych.

## Trening kontrastujący (MLCL)

Proponowane rozszerzenie funkcji Supervised Contrastive Loss [Khosla et al., 2020] do obserwacji z wieloma etykietami:

$$\mathcal{L}^{\text{mlc}} = \sum_{i=1}^{2N} \mathcal{L}_i^{\text{mlc}} \quad (5)$$

## Trening kontrastujący (MLCL)

Proponowane rozszerzenie funkcji Supervised Contrastive Loss [Khosla et al., 2020] do obserwacji z wieloma etykietami:

$$\mathcal{L}^{\text{mlc}} = \sum_{i=1}^{2N} \mathcal{L}_i^{\text{mlc}} \quad (5)$$

$$\mathcal{L}_i^{\text{mlc}} = \frac{1}{2N_{\tilde{Y}_i} - 1} \sum_{j=1}^{2N} \mathbb{1}_{i \neq j} \cdot \mathbb{1}_{\tilde{Y}_i \cap \tilde{Y}_j \neq \emptyset} \cdot \mathcal{L}_{i,j}^{\text{mlc}} \quad (6)$$

## Trening kontrastujący (MLCL)

Proponowane rozszerzenie funkcji Supervised Contrastive Loss [Khosla et al., 2020] do obserwacji z wieloma etykietami:

$$\mathcal{L}^{\text{mlc}} = \sum_{i=1}^{2N} \mathcal{L}_i^{\text{mlc}} \quad (5)$$

$$\mathcal{L}_i^{\text{mlc}} = \frac{1}{2N_{\tilde{Y}_i} - 1} \sum_{j=1}^{2N} \mathbb{1}_{i \neq j} \cdot \mathbb{1}_{\tilde{Y}_i \cap \tilde{Y}_j \neq \emptyset} \cdot \mathcal{L}_{i,j}^{\text{mlc}} \quad (6)$$

$$\mathcal{L}_{i,j}^{\text{mlc}} = -\log \frac{\exp(z_i \cdot z_j / \tau)}{\sum_{k=1}^{2N} \mathbb{1}_{i \neq k} \cdot \exp(z_i \cdot z_k / \tau) + \mathcal{L}_{i,j,k}^{\text{mlc}}} \quad (7)$$

## Trening kontrastujący (MLCL)

Proponowane rozszerzenie funkcji Supervised Contrastive Loss [Khosla et al., 2020] do obserwacji z wieloma etykietami:

$$\mathcal{L}^{\text{mlc}} = \sum_{i=1}^{2N} \mathcal{L}_i^{\text{mlc}} \quad (5)$$

$$\mathcal{L}_i^{\text{mlc}} = \frac{1}{2N_{\tilde{Y}_i} - 1} \sum_{j=1}^{2N} \mathbb{1}_{i \neq j} \cdot \mathbb{1}_{\tilde{Y}_i \cap \tilde{Y}_j \neq \emptyset} \cdot \mathcal{L}_{i,j}^{\text{mlc}} \quad (6)$$

$$\mathcal{L}_{i,j}^{\text{mlc}} = -\log \frac{\exp(z_i \cdot z_j / \tau)}{\sum_{k=1}^{2N} \mathbb{1}_{i \neq k} \cdot \exp(z_i \cdot z_k / \tau) + \mathcal{L}_{i,j,k}^{\text{mlc}}} \quad (7)$$

$$\mathcal{L}_{i,j,k}^{\text{mlc}} = \sum_{l=1}^7 \exp(z_i \cdot z'_{k,l} / \tau) \quad (8)$$

## Trening kontrastujący (MLCL)

Proponowana funkcja łączona jest z treningiem pomocniczym z kodowaniem rzadkim:

$$\mathcal{L} = \mathcal{L}^{\text{mlc}} + \beta \mathcal{L}^{\text{aux}} \quad (9)$$

gdzie  $\beta$  określa kompromis pomiędzy nauką poprzez kontrastowanie oraz treningiem pomocniczym.

Analiza gradientu wykazała, że funkcja  $\mathcal{L}^{\text{mlc}}$ :

- skupia się na trudnych obserwacjach,
- nie skupia się na prostych obserwacjach.

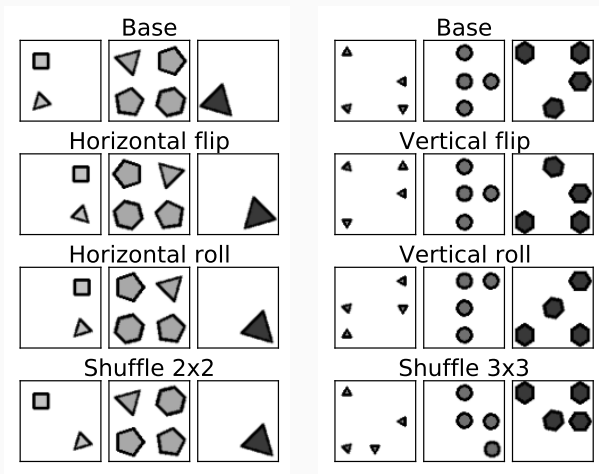
Pozwala to uniknąć kosztownych operacji wyszukiwania trudnych obserwacji (ang. *hard example mining*).



## Trening kontrastujący (MLCL)

Metoda		Skuteczność (%)
		$\mu \pm \sigma$
SCL	CE	$82.8 \pm 0.7$
	AUX-sparse	$95.6 \pm 0.5$
	MLCL	<b><math>95.7 \pm 0.5</math></b>
SRAN	CE	$56.0 \pm 1.8$
	AUX-sparse	$59.9 \pm 1.6$
	MLCL	<b><math>66.1 \pm 1.6</math></b>
CoPINet	CE	$44.8 \pm 0.8$
	AUX-sparse	$50.3 \pm 1.0$
	MLCL	<b><math>52.3 \pm 1.1</math></b>

**Tablica 2:** Skuteczność testowa modeli trenowanych poprzez kontrastowanie na zbiorze danych I-RAVEN uśredniona z 4 przebiegów z losową inicjalizacją.



**Rysunek 34:** Augmentacja RPM ze zbioru danych I-RAVEN. Wybrana transformacja jest stosowana w ten sam sposób do wszystkich paneli danej matrycy.

Metoda	DA	Skuteczność (%)	
		CE	
SCL	×	82.8	
	✓	90.8	
SRAN	×	56.0	
	✓	65.8	
CoPINet	×	44.8	
	✓	48.3	

**Tablica 3:** Wpływ wzbogacenia danych na skuteczność testową modeli trenowanych na zbiorze danych I-RAVEN uśredniona z 4 przebiegów z losową inicjalizacją.

Metoda	DA	Skuteczność (%)	
		CE	AUX-dense
SCL	×	82.8	86.9
	✓	90.8	95.3
SRAN	×	56.0	59.4
	✓	65.8	69.1
CoPINet	×	44.8	32.2
	✓	48.3	35.2

**Tablica 3:** Wpływ wzbogacenia danych na skuteczność testową modeli trenowanych na zbiorze danych I-RAVEN uśredniona z 4 przebiegów z losową inicjalizacją.

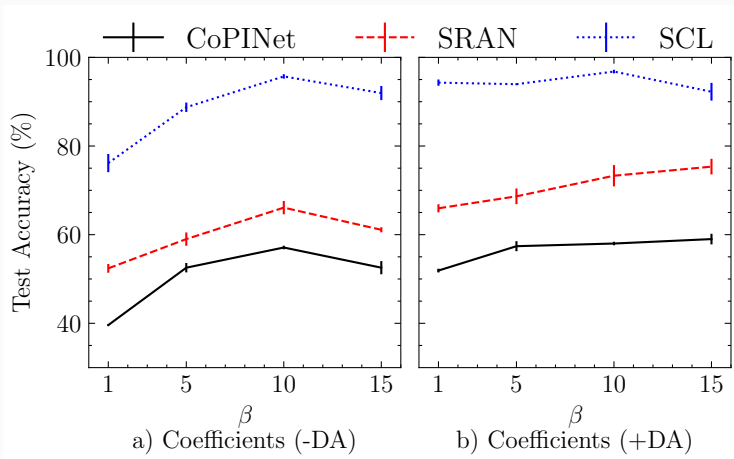
Metoda	DA	Skuteczność (%)		
		CE	AUX-dense	AUX-sparse
SCL	×	82.8	86.9	95.6
	✓	90.8	95.3	95.9
SRAN	×	56.0	59.4	59.9
	✓	65.8	69.1	72.2
CoPINet	×	44.8	32.2	50.3
	✓	48.3	35.2	49.2

**Tablica 3:** Wpływ wzbogacenia danych na skuteczność testową modeli trenowanych na zbiorze danych I-RAVEN uśredniona z 4 przebiegów z losową inicjalizacją.

Metoda	DA	Skuteczność (%)			
		CE	AUX-dense	AUX-sparse	MLCL
SCL	×	82.8	86.9	95.6	95.7
	✓	90.8	95.3	95.9	<b>96.8</b>
SRAN	×	56.0	59.4	59.9	66.1
	✓	65.8	69.1	72.2	<b>73.3</b>
CoPINet	×	44.8	32.2	50.3	52.3
	✓	48.3	35.2	49.2	<b>57.1</b>

**Tablica 3:** Wpływ wzbogacenia danych na skuteczność testową modeli trenowanych na zbiorze danych I-RAVEN uśredniona z 4 przebiegów z losową inicjalizacją.

# Kompromis treningu pomocniczego i kontrastującego



**Rysunek 35:** Badania ablacyjne metody MLCL na zbiorze danych I-RAVEN, pokazujące kompromis pomiędzy treningiem pomocniczym i kontrastującym.

## Elementy składowe funkcji kosztu

Model	$\mathcal{L}$	Skuteczność (%)
SCL	95.7	
SRAN	59.4	
CoPINet	52.3	

**Tablica 4:** Badania ablacyjne metody MLCL na zbiorze danych I-RAVEN, pokazujące wpływ elementów składowych funkcji kosztu.



## Elementy składowe funkcji kosztu

Model	Skuteczność (%)	
	$\mathcal{L}$	bez $\mathcal{L}_{i,j,k}^{\text{mlc}}$
SCL	95.7	66.3
SRAN	59.4	31.5
CoPINet	52.3	26.8

**Tablica 4:** Badania ablacyjne metody MLCL na zbiorze danych I-RAVEN, pokazujące wpływ elementów składowych funkcji kosztu.

## Elementy składowe funkcji kosztu

Model	Skuteczność (%)		
	$\mathcal{L}$	bez $\mathcal{L}_{i,j,k}^{\text{mlc}}$	$\mathcal{L}^{\text{aux}}$
SCL	95.7	66.3	55.1
SRAN	59.4	31.5	50.0
CoPINet	52.3	26.8	36.5

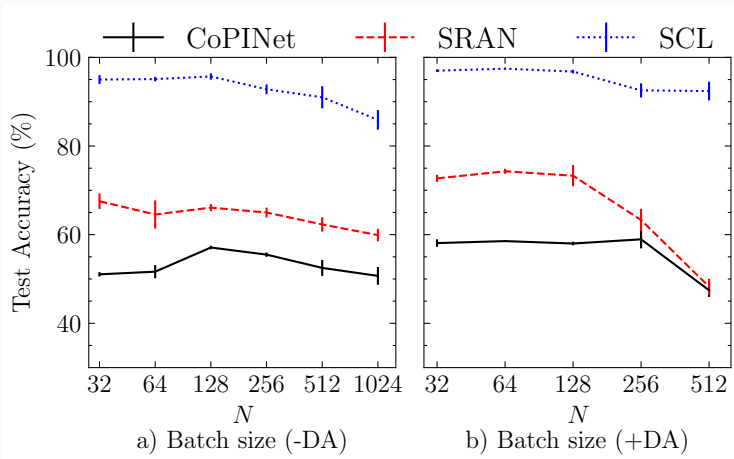
**Tablica 4:** Badania ablacyjne metody MLCL na zbiorze danych I-RAVEN, pokazujące wpływ elementów składowych funkcji kosztu.

## Elementy składowe funkcji kosztu

Model	Skuteczność (%)			
	$\mathcal{L}$	bez $\mathcal{L}_{i,j,k}^{\text{mlc}}$	$\mathcal{L}^{\text{aux}}$	$\mathcal{L}^{\text{mlc}}$
SCL	95.7	66.3	55.1	54.2
SRAN	59.4	31.5	50.0	29.0
CoPINet	52.3	26.8	36.5	26.5

**Tablica 4:** Badania ablacyjne metody MLCL na zbiorze danych I-RAVEN, pokazujące wpływ elementów składowych funkcji kosztu.

# Rozmiar batcha



**Rysunek 36:** Badania ablacyjne metody MLCL na zbiorze danych I-RAVEN, pokazujące wpływ rozmiaru batcha.

## Badania generalizacji na zbiorze PGM

Reżim	Skuteczność (%)			
	CE	AUX-dense	AUX-sparse	MLCL
Neutral	85.6	<b>87.1</b>	<b>87.1</b>	71.1
Interpolation	55.8	56.0	54.1	<b>70.9</b>
H.O. Attribute Pairs	40.8	<b>79.6</b>	63.6	66.0
H.O. Triple Pairs	64.5	<b>76.6</b>	64.0	71.7
H.O. Triples	27.0	23.0	<b>30.8</b>	22.1
H.O. Line-Type	15.1	14.1	<b>17.0</b>	16.1
H.O. Shape-Colour	12.7	12.6	12.7	12.8
Extrapolation	17.3	19.8	17.5	<b>21.9</b>

**Tablica 5:** Skuteczność modelu SCL we wszystkich reżimach zbioru danych PGM.

# Podsumowanie

---

- Dobór możliwych odpowiedzi może mieć duże znaczenie
- Nauka nadzorowana poprzez kontrastowanie może być alternatywą do cross-entropy
- Metody kontrastowe mogą wymagać dużych rozmiarów batcha

Zapraszam do dyskusji



## Literatura

---

David G. T. Barrett, Felix Hill, Adam Santoro, Ari S. Morcos, and Timothy P. Lillicrap. Measuring abstract reasoning in neural networks. *International Conference on Machine Learning (ICML)*, 2018.

Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. *arXiv preprint arXiv:2002.05709*, 2020.

Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. *Advances in neural information processing systems*, 27:2672–2680, 2014.

Kaiming He, Haoqi Fan, Yuxin Wu, Saining Xie, and Ross Girshick. Momentum contrast for unsupervised visual representation learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9729–9738, 2020.

Felix Hill, Adam Santoro, David Barrett, Ari Morcos, and Timothy Lillicrap. Learning to Make Analogies by Contrasting Abstract Relational Structure. In *International Conference on Learning Representations*, 2019.

- Prannay Khosla, Piotr Teterwak, Chen Wang, Aaron Sarna, Yonglong Tian, Phillip Isola, Aaron Maschiot, Ce Liu, and Dilip Krishnan. Supervised Contrastive Learning, 2020.
- Jacek Mańdziuk and Adam Żychowski. DeepIQ: A Human-Inspired AI System for Solving IQ Test Problems. In *2019 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8. IEEE, 2019.
- David Silver, Julian Schrittwieser, Karen Simonyan, Ioannis Antonoglou, Aja Huang, Arthur Guez, Thomas Hubert, Lucas Baker, Matthew Lai, Adrian Bolton, et al. Mastering the game of go without human knowledge. *nature*, 550(7676):354–359, 2017.

Tongzhou Wang and Phillip Isola. Understanding Contrastive Representation Learning through Alignment and Uniformity on the Hypersphere. *arXiv preprint arXiv:2005.10242*, 2020.

Chi Zhang, Feng Gao, Baoxiong Jia, Yixin Zhu, and Song-Chun Zhu. Raven: A dataset for relational and analogical visual reasoning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5317–5327, 2019.