



POZNAN UNIVERSITY OF TECHNOLOGY

FACULTY OF COMPUTING AND TELECOMMUNICATION
Institute of Computing Science

PhD Thesis

STAGED SELF-SUPERVISION FOR SOLVING RAVEN'S PROGRESSIVE MATRICES

Jakub Kwiatkowski

Supervisor
prof. dr hab. inż. Krzysztof Krawiec

POZNAN 2026

Abstract

Solving visual reasoning tasks requires more than recognizing surface patterns: it demands the capacity to discover and apply abstract relational rules across novel configurations. Raven’s Progressive Matrices (RPMs)—visual puzzles in which a missing panel must be inferred to complete a structured geometric pattern—have become a standard benchmark for assessing such capabilities in machine learning systems. Contemporary approaches predominantly rely on monolithic architectures trained to select the correct answer from a fixed candidate pool. This formulation conflates perception, reasoning, and decision-making into a single undifferentiated process, limiting the model’s ability to acquire structured representations of the underlying rules and creating systematic vulnerabilities to statistical regularities in the answer set—a class of failure modes known as shortcut learning.

This thesis proposes a staged self-supervision framework that reformulates RPM solving by decomposing it into two distinct stages: attribute prediction and choice making. The first stage employs the Abstract Compositional Transformer (ACT), a transformer-based architecture that learns to infer structured, symbolic descriptions of RPM panels by predicting masked entries from the surrounding context. The second stage introduces three alternative choice-making modules—the Direct Choice Maker, the Learned Choice Maker, and the Contrastive Choice Maker—each comparing the predicted attributes with those of answer panels to identify the correct solution.

The framework achieves results competitive with the strongest approaches reported on the RAVEN, I-RAVEN, and RAVEN-FAIR benchmarks, without requiring dataset-specific retraining for each. The staged decomposition additionally mitigates dataset biases by architectural design: the attribute prediction stage operates exclusively on context panels without answer set exposure, while each choice-making module evaluates candidates independently, eliminating pathways for statistical exploitation. Extensive empirical analysis—including ablation studies, data efficiency experiments, and visualization of learned representations—confirms that the framework acquires compositional structure rather than surface-level correlations. The work establishes that principled decomposition combined with self-supervised learning enables neural networks to discover structured representations of abstract visual rules.

Keywords: abstract reasoning, Raven’s Progressive Matrices, self-supervised learning, transformer architectures, shortcut learning, representation learning

Streszczenie

Skuteczne rozwiązywanie zadań wymagających rozumowania wizualnego nie polega jedynie na rozpoznawaniu powierzchownych wzorców. Wymaga uchwycenia abstrakcyjnych zależności między elementami oraz zastosowania ich w nowych konfiguracjach. Progresywne Matryce Ravena (RPM) — wizualne zagadki, w których należy wskazać brakujący panel uzupełniający wzorec — stały się powszechnie stosowanym benchmarkiem służącym do oceny takich zdolności w systemach uczenia maszynowego. Dominujące podejścia opierają się na monolitycznych architekturach uczonych do wskazywania poprawnej odpowiedzi spośród ustalonego zbioru kandydatów. Takie ujęcie zadania zaciera granice między percepcją, rozumowaniem i podejmowaniem decyzji, przez co utrudnia modelowi nabywanie reprezentacji reguł rządzących wzorcem. Jednocześnie sprzyja wykorzystywaniu statystycznych regularności obecnych w zbiorze odpowiedzi zamiast rzeczywistego rozumowania — zjawisko to określa się mianem uczenia na skrótach (*shortcut learning*).

Niniejsza praca proponuje metodę samonadzorowanego uczenia etapowego (*staged self-supervision*), która przeformułowuje problem rozwiązywania RPM jako dwuetapowy proces obejmujący predykcję atrybutów oraz wybór odpowiedzi. W pierwszym etapie zastosowano Abstract Compositional Transformer (ACT) — architekturę transformerową uczącą się przewidywać symboliczne opisy paneli RPM poprzez rekonstrukcję zamaskowanych paneli na podstawie otaczającego kontekstu. Drugi etap realizują trzy alternatywne moduły wyboru odpowiedzi: *Direct Choice Maker*, *Learned Choice Maker* oraz *Contrastive Choice Maker*, z których każdy porównuje przewidywane atrybuty z atrybutami paneli kandydujących i wskazuje rozwiązanie.

Zaproponowana metoda osiąga wyniki konkurencyjne wobec najlepszych opublikowanych rozwiązań na benchmarkach RAVEN, I-RAVEN oraz RAVEN-FAIR, bez konieczności odrębnego trenowania dla każdego z nich. Dwuetapowy proces ogranicza ponadto wpływ obciążeń zbioru danych już na poziomie architektury: etap predykcji atrybutów operuje wyłącznie na panelach kontekstowych, natomiast każdy moduł ocenia kandydatów niezależnie, uniemożliwiając eksploatację statystycznych regularności zbioru odpowiedzi. Obszerna analiza empiryczna — obejmująca badania ablacyjne, eksperymenty dotyczące efektywności danych oraz wizualizację nauczonych reprezentacji — potwierdza, że metoda uczy się reguł rządzących wzorcem, a nie jedynie powierzchownych zależności statystycznych. Wyniki wskazują, że prze-myślana dekompozycja zadania połączona z samonadzorowanym uczeniem pozwala sieciom neuronowym

skutecznie rozwiązywać zadania abstrakcyjnego rozumowania.

Słowa kluczowe: rozumowanie abstrakcyjne, Progresywne Matryce Ravena, samonadzorowane uczenie, architektury transformerowe, uczenie na skrótach (*shortcut learning*), uczenie reprezentacji