

## 第7章 ARC-AGIを拡散モデルで解けないかな

著：@kyomu

### 7.1 はじめに

今日、生成AIの目覚ましい発展により、私たちの勉強への向き合い方、労働のあり方などは急速に変容しつつあります。特に、ソフトウェアエンジニアリング領域における生成AIの影響は非常に大きいです。Claude Codeをはじめとするコーディングエージェントはコーディングの障壁を低くし、今までにない速度でのアプリケーション開発を実現させました。

そんな生成AIですが、最近のモデルは難しい数学の問題や、競技プログラミングの問題を解けるようになってきており、生成AIがとてつもない知能を有しているように見えます。しかし、それは本当に**知能**と言えるのでしょうか。

この世界には、数多くの論文、数学の参考書、プログラミングコンテストで出題された過去問など大量の情報源が存在しています。たとえ生成AIが問題を解けたとしても、その問題が過去に問題集で取り上げられたものと類似していれば、AIは知能を用いて問題を解いたというよりは暗記した解法で解いたというのがふさわしいように思えます。

すると、生成AIが真の知能、ここでは**未知の課題に対する適応力(流動性知能)**を有しているかが興味の対象となります。そこで、生成AIが流動性知能を有しているかを測るベンチマークとしてARC-AGIが考案されました。後述するように、ARC-AGIは人間にとっては非常に簡単なタスクですが、生成AIにとっては非常に難易度の高いタスクとなっています。

本章では、ARC-AGIの解法にQwenといったLLMを使用しないアプローチとして拡散モデルを用いることができないかというアイデアについて話します。

### 7.2 ARC-AGIについて

ARC-AGIとはAIの流動性知能を測定するベンチマークの1つです。2次元グリッドを入力として受け取り、正解となる2次元グリッドを出力するようなタスクを解きます。このとき、1つから5つの入出力例が与えられるので、入出力間のグリッドの変換規則を理解し、新たな入力に対応できるような能力が求められます。

#### 具体的なタスク

具体的なタスクを解いてみることで、ARC-AGIの難易度感を把握してみましょう。ここではARC-AGI-1 Public TrainingのID: `1e32b0e9` を取り上げます。図1の入出力例からパターンを見出し、図2から期待される出力を当ててみましょう。

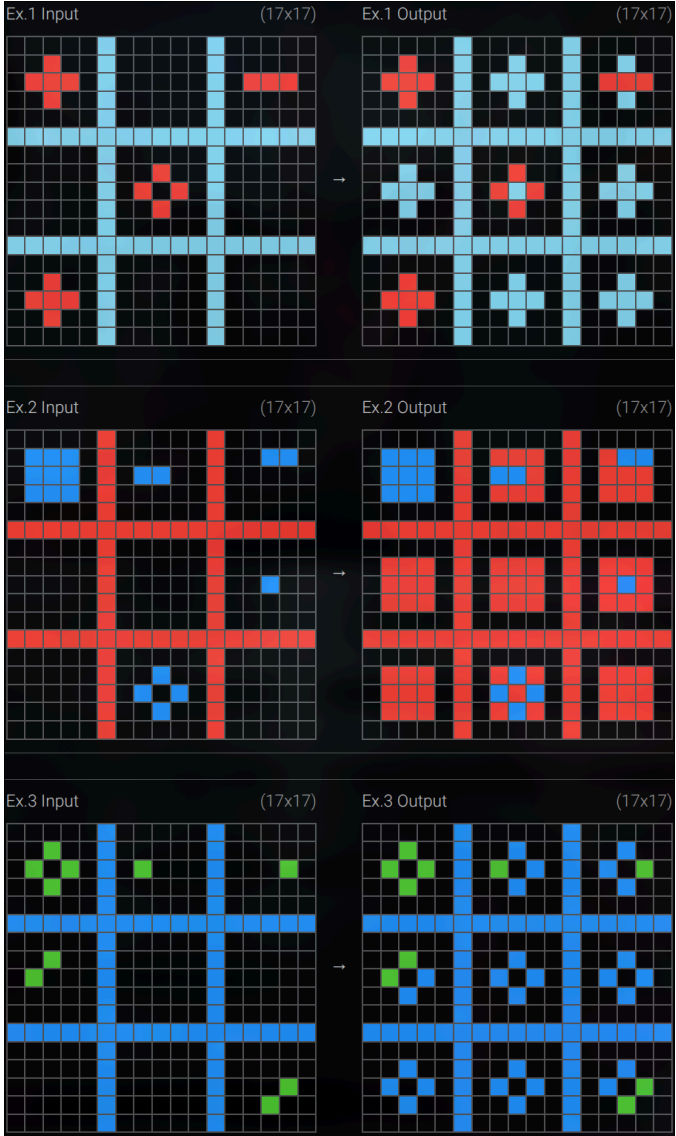


図1: 入出力例

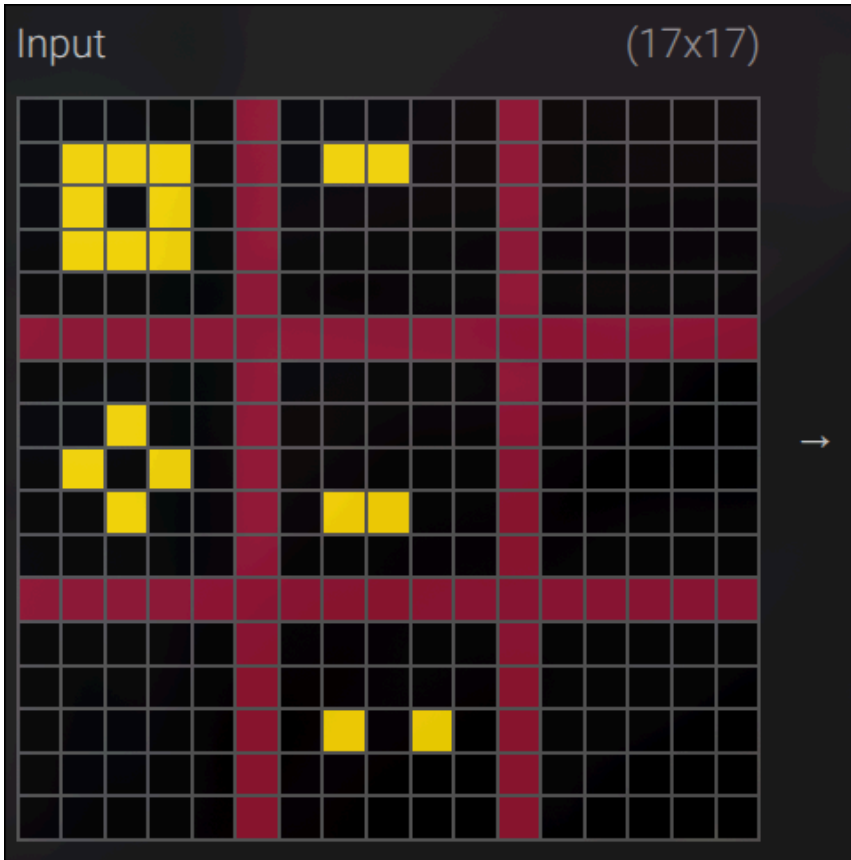


図2: 問題

解けたでしょうか。このように、ARC-AGIは、特定の変換規則さえ見出すことができれば簡単に解くことが可能なタスクとなっています。

ARC-AGIは、人間にとって極めて難しいタスクではないと思われます。しかし、生成AIにとってはどうでしょうか。ARC-AGI-2ベンチマークに対する、各モデルの正解率(2026年2月21日時点)を見てみましょう。

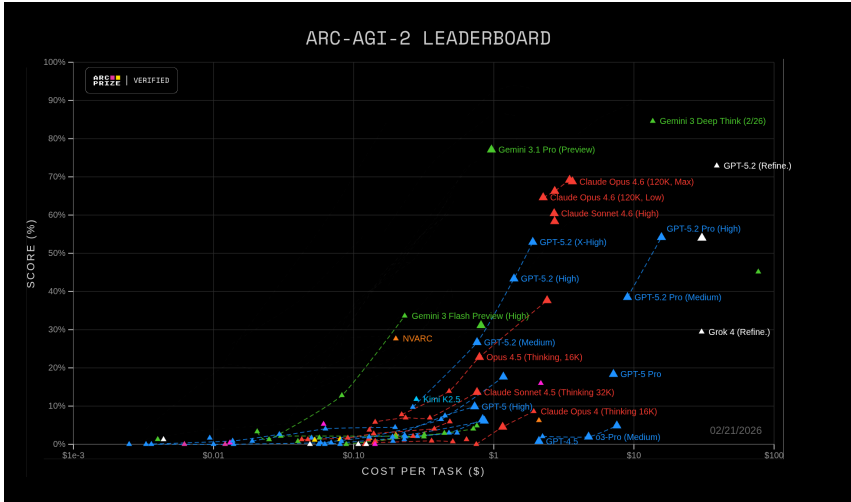


図3: ARC-AGI-2 leaderboard(横軸は1タスクあたりのコスト、縦軸は正答率を表す。)

GPT-4.5等の強力なモデルでさえも、正答率が10%を下回るというのが注目すべき点です。人間はARC-AGI-2のタスクを60%ほどの正答率で解くことができることを考えると、GPT-4.5が登場した頃は、まだまだAIが流動性性能を有していなかったと言えるでしょう。ここ最近になって、Gemini 3やGPT-5.2といった高い正答率を持つモデルが現れるようになったものの、いずれもCoT<sup>\*1</sup>による高い推論コスト(1タスクあたりに\$1~\$100)を要します。また、ARC-AGI-1, ARC-AGI-2ともに正答率100%を出せるモデルは未だに登場していません。

### 7.3 拡散モデルを用いる動機

ARC-AGIの各タスクを、短時間かつ低コストで解くことができるモデルを構築することを考えます。

Kaggle<sup>\*2</sup>で開催されたARC-Prizeというコンペティションでは、ARC-AGIベンチマークを限られた計算資源の下で高速に解くことのできるモデルを構築することが要求されました。ARC-Prizeの上位解法には、Qwen<sup>\*3</sup>といったLLMが使用されているという特徴があり、これが基盤モデルを用いない解法を考えるきっかけとなりました。

1. Chain of Thought(思考の連鎖)の略。AIに最終的な答えを出す前の論理的な推論過程を段階的に出力させることで、複雑な問題に対する回答の精度を飛躍的に向上させる手法のことです。
2. Kaggleとは、世界中のデータサイエンティストがAIモデルの予測精度を競い合う、世界最大の機械学習プラットフォームです。
3. Qwen等の軽量なモデルが採用されたのはKaggleにおけるコンペティションの制約のためです。

## 対称性へのアプローチ

ある日、拡散モデルの書籍を読み進めていると、次のような記述を見つけました。

拡散モデルは生成における対称性をモデルに組み込むことができる。

\*4

創薬などの化合物生成の分野では、拡散モデルによって良い結果が得られたという報告を度々目にします。この対称的な構造を上手く扱えるという特徴は、ARC-AGIにとって有効なものであると考えました。というのも、ARC-Prizeの訓練データには同じ構造の繰り返し、部分的な回転・鏡像対称性をもつタスクがいくつか確認されたからです。

## 柔軟な条件付け

ARC-AGIの各タスクは、入力という条件の下で、目標のグリッドを生成する条件付き生成の問題であると考えられます。拡散モデルでは、SBM<sup>\*5</sup>を考えた際の各 $t$ でのスコア $\nabla \log p_t(x)$ を条件付きスコア $\nabla \log p_t(x|y)$ に変えるだけで条件付き生成を実現できます。拡散モデルは自然に条件付き生成を導入できるというのが魅力であり、以上のようにして拡散モデルを用いてARC-AGIを解くことを考えるようになりました。

## 7.4 離散拡散モデルを導入する

### 連続空間では不適切

拡散モデルを直接用いるアプローチでは問題がありました。ARC-AGIの各タスクにおけるグリッドは、0 から 9 の整数値で構成されています。

4. 岡野原大輔, 拡散モデル - データ生成技術の数理. 岩波書店, 2023.

5. Score-Based generative Model(スコアベース生成モデル)の略。確率密度関数の対数微分 $\nabla \log p(x)$ はスコア関数と呼ばれ、この関数を学習して元のデータ分布を推定・生成する手法のこと。DDPMなどの拡散モデルにおける、ノイズを予測する過程は、ノイズ付きのデータ分布のスコアを推定することと等価であることが示せる。

```

display(task['train'][0]['input'])
display(task['train'][0]['output'])

[[6, 6, 0], [6, 0, 0], [0, 6, 6]]

[[6, 6, 0, 6, 6, 0, 0, 0, 0],
 [6, 0, 0, 6, 0, 0, 0, 0, 0],
 [0, 6, 6, 0, 6, 6, 0, 0, 0],
 [6, 6, 0, 0, 0, 0, 0, 0, 0],
 [6, 0, 0, 0, 0, 0, 0, 0, 0],
 [0, 6, 6, 0, 0, 0, 0, 0, 0],
 [0, 0, 0, 6, 6, 0, 6, 6, 0],
 [0, 0, 0, 6, 0, 0, 6, 0, 0],
 [0, 0, 0, 0, 6, 6, 0, 6, 6]]

```

図4: 入出力例に対応するグリッド

よって、グリッドの色自体には意味が含まれていないことに注意が必要です。ARC-AGIで用いるグリッドは、現実の写真が性質として持つピクセル間の滑らかな移り変わりが存在していません。このような問題設定の下で、連続的な実数を扱う通常の拡散モデルを用いるのは適していないように考えられます。

そこで離散拡散モデル D3PM<sup>\*6</sup>を用います。拡散モデルというと、拡散過程でデータへノイズを付加し、逆拡散過程でノイズを予測するという流れで学習します。しかし、離散値やカテゴリカル

データに対して、ノイズ付加は単なる数値加算によって実現できません。この問題に対して、離散拡散モデルでは、遷移行列を用いたトークンの置き換えによって従来の拡散モデルにおけるノイズ付加を実現することで対応します。

## 固定長の入出力へ

ARC-AGIでは可変長の入出力が可能なモデルを求められています。ですが、可変長の入出力は扱いにくいので、固定長のグリッドを扱う問題として考えられないかを試みます。

入出力で扱うグリッドの大きさには上限があり、 $32 \times 32$  が最大のサイズとのことでした。ここでは思い切って全ての入出力を  $32 \times 32$  で行うことについて検討します。元々、グリッドの各ピクセルを表す整数値は 0 から 9 ですが、ここに 10 という特別な整数値を導入します。10 は「区切り」を表す整数値とし、元の入力のグリッドを次のように区切りを挟んで繰り返した  $32 \times 32$  サイズのグリッドを作成します。



図5: 元の入力

6. Austin et al. (2023) Structured Denoising Diffusion Models in Discrete State-Spaces.

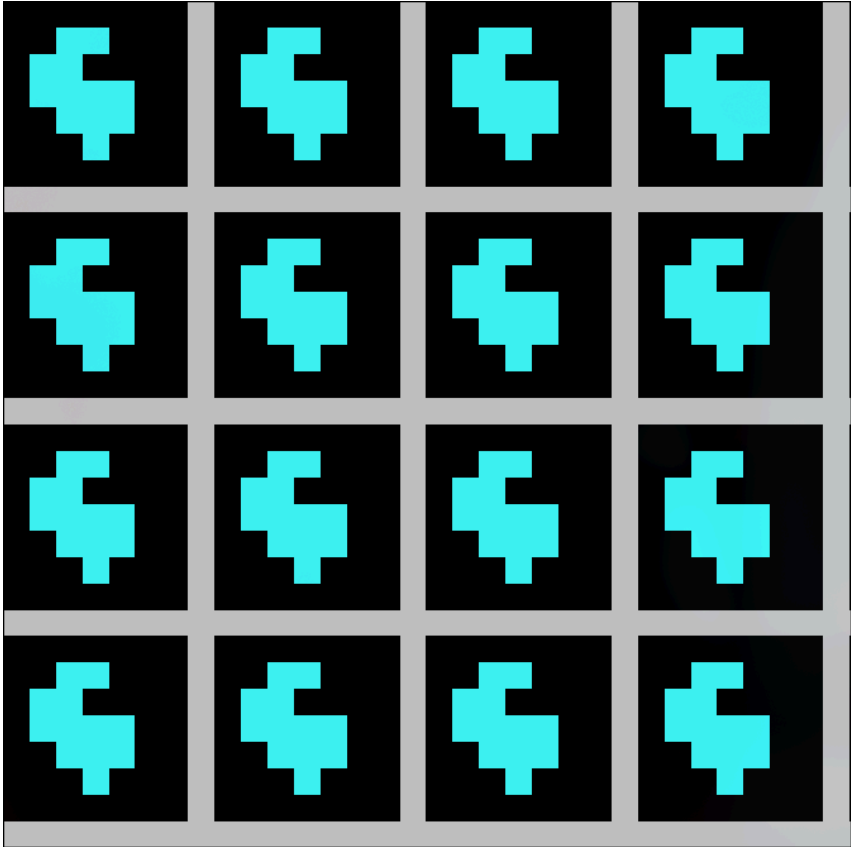
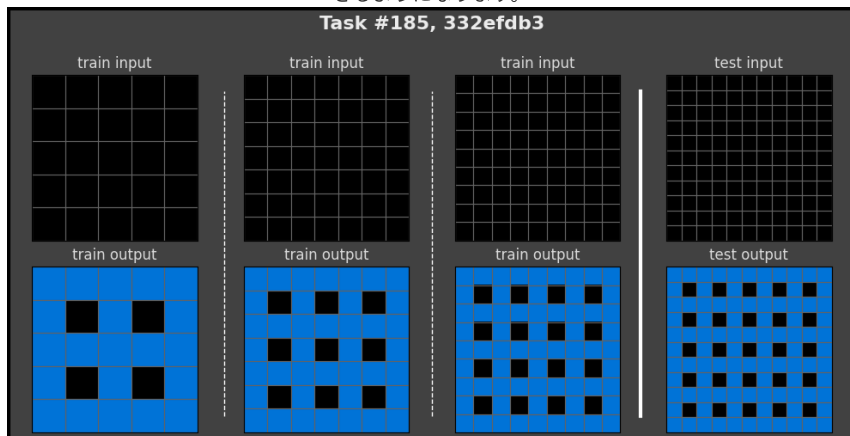


図6: 32×32サイズに変形させた入力

十分に良い品質の出力が得られると仮定すれば、この手法で出力された 32×32 サイズのグリッドから求めた出力を切り抜くことができます。

また、このように区切りをもとにグリッドを繰り返す手法には、グリッドのサイズの情報を自然に与えることができるというメリットもあります。

下図のようなタスクでは、グリッドのサイズの情報がなく、出力のグリッドのサイズを決定できません。区切りを導入することで、区切りの間隔を手がかりにグリッドのサイズをモデルは把握できるようにします。



7

## 遷移行列の設計

D3PMの拡散過程に用いられる遷移行列にはいくつかの種類が考案されています。ここでは、そのうちの吸収状態(原文: Absorbing state)と呼ばれる遷移行列を用いる場合を考えます。

### サンプリングの高速化

拡散過程において、ある時刻 $t$ でのデータ $x_t$ は、遷移行列 $Q_t$ を用いて、以下の確率分布から生成されます。

$$q(x_t|x_{t-1}) = x_{t-1}Q_t$$

この定義では、 $x_t$ を計算するには、 $x_{t-1}$ の値を参照する必要があります。よって、任意の時刻 $t$ でのデータ $x_t$ は、初期状態 $x_0$ から計算するのに $O(t)$ の時間がかかります。これでは学習が非効率なので、累積遷移行列 $\bar{Q}_t$ を用いて次のようにサンプリングを行います。

$$\begin{aligned}\bar{Q}_t &= Q_1Q_2\cdots Q_t \\ q(x_t|x_0) &= x_0\bar{Q}_t\end{aligned}$$

このように、前計算による $O(1)$ へのサンプリング高速化を実現します。

### 吸収状態の遷移行列

吸収状態の遷移行列は、ある [MASK] と呼ばれる特定の状態に遷移した後は、そこから動かなくなるような遷移行列です。詳細には、添字 $i, j$ をカテゴリ (ARC-AGIでは各ピクセルを表す整数値)

とし、**[MASK]**を表す添字を $m$ とすると、吸収状態の遷移行列は次のように定義されます。

$$[Q_t]_{ij} = \begin{cases} 1 & i = j = m \\ 1 - \beta & i = j \neq m \\ \beta & i \neq j, j = m \end{cases}$$

ここで $\beta$ は時刻 $t$ によって変化するノイズスケジューラで、0から1の値をとります。

吸収状態の遷移行列を用いた場合、初期状態 $x_0$ から、ある時刻 $t$ でのデータ $x_t$ はより簡単に計算できます。

具体的には、 $\bar{\beta}_t = \prod_{s=1}^t (1 - \beta_s)$ とおくと、初期状態 $x_0$ の各トークンは、次のようにして $x_t$ の各トークンに変換されます。

- 確率 $1 - \bar{\beta}_t$ で、 $x_0$ のトークンが**[MASK]**に置き換わる。
- 確率 $\bar{\beta}_t$ で、 $x_0$ のトークンがそのまま残る。

このように、吸収状態の遷移行列を用いると、複雑な行列計算を行うことなく、初期状態 $x_0$ から任意の時刻 $t$ でのデータ $x_t$ を簡単に計算できます。

計算効率以外の面でも、吸収状態の遷移行列を使うことには理由があります。ARC-AGIのタスクの一部には、下図のように欠損したグリッドを特定の規則で補完させるようなものが確認されました。

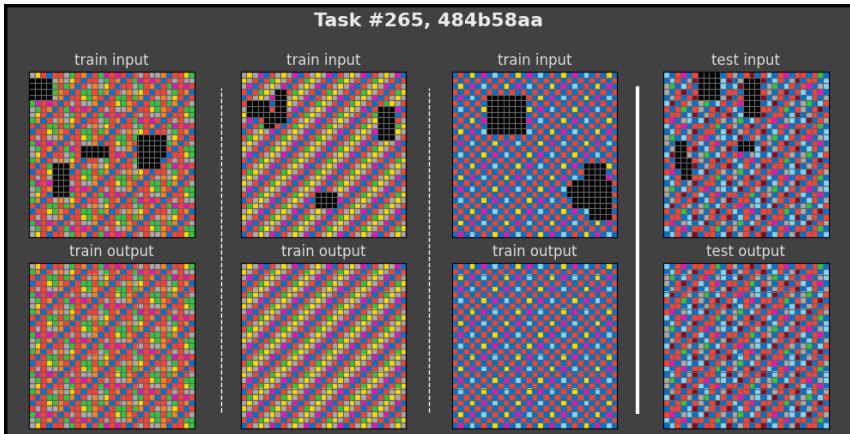


図7: 欠損したグリッドを補完するタスクの例

吸収状態の遷移行列を用いた場合、逆拡散過程で行う処理は、**[MASK]**に置き換わったトークンを、周りのトークンを手がかりに正しいトークンへ置き換えるというものになります。このような処理は、欠損したグリッドを補完するタスクと類似しているため、自然に補完の能力をモデルに組み込めることを期待できます。

## 7.5 結果

残念ながら現時点で、離散拡散モデルを使用した場合、連続拡散モデルを使用した場合ともに良い性能は確認できていません。

連続拡散モデルで、ある評価データに対する出力を確認すると、下図のような結果が得られました。

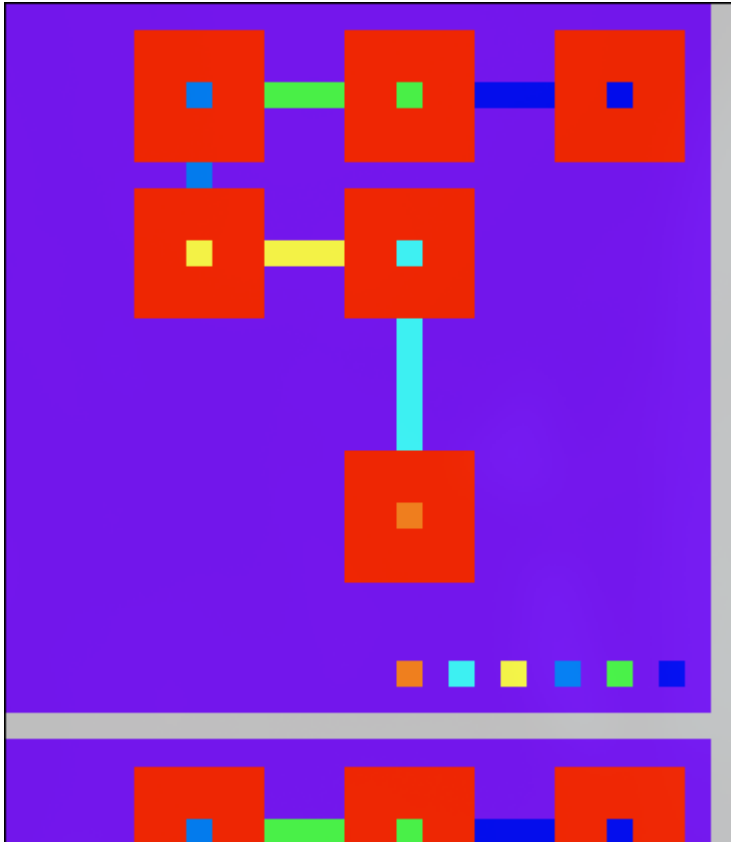


図8: 期待される出力

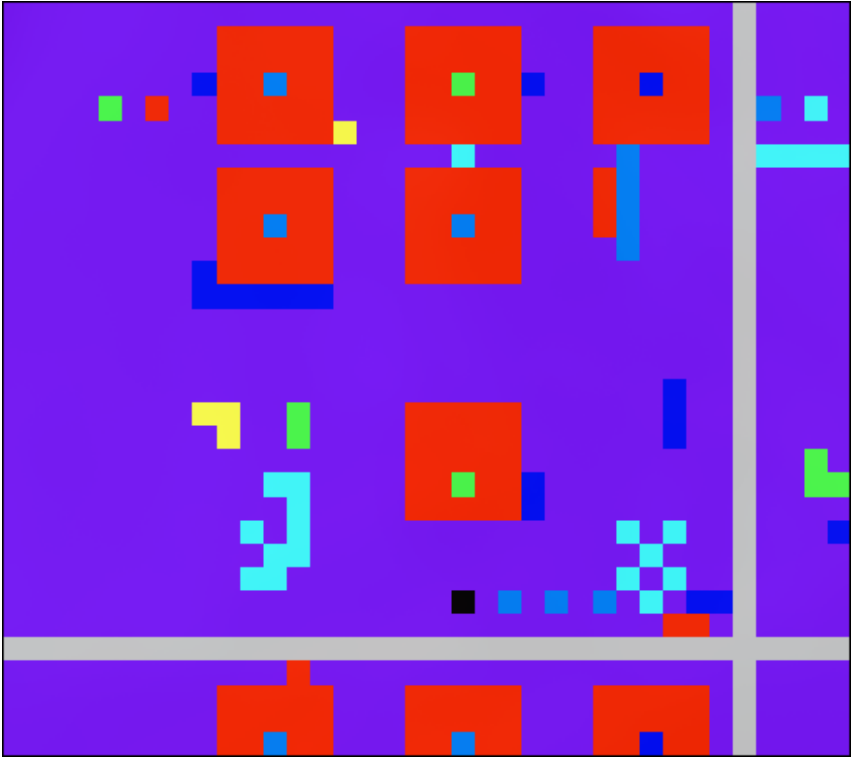


図9: モデルの出力

かなり近い構造をモデルは出力できているものの、まだまだタスクに正答するのは難しそうです。

## 7.6 おわりに

お読みいただきありがとうございました。ARC Prize Foundation<sup>\*7</sup>で掲げられているように、新しいアイデアがARC-AGIを解くには求められています。もし興味を持っていただけたら、皆さんも挑戦してみてください。

7. <https://arcprize.org/about>