1. Preliminary on Hyperplane Arrangements

In this section, we recall some basic knowledge on hyperplane arrangements (Zaslavsky, 1975; Stanley, 2004), which will be used in the proofs of theorems in this paper. An affine hyperplane in a Euclidean space \( V \simeq \mathbb{R}^n \) is a subspace with the following form: \( H = \{ X \in V : \alpha \cdot X = b \} \), where \( \cdot \) denotes the inner product, \( 0 \neq \alpha \in V \) is called the norm vector of \( H \), and \( b \in \mathbb{R} \). For example, when \( V = \mathbb{R}^n \), an affine hyperplane has the following form: \( \{ (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n : \sum_{i=1}^{n} a_i x_i = b \} \) where \( a_i, b \in \mathbb{R} \) and there exists some \( i \) with \( a_i \neq 0 \). A finite hyperplane arrangement \( \mathcal{A} \) of a Euclidean space \( V \) is a finite set of affine hyperplanes in \( V \). A region of an arrangement \( \mathcal{A} = \{ H_i \subset V : 1 \leq i \leq m \} \) is defined as a connected component of \( V \setminus (\bigcup_{i=1}^{m} H_i) \), which is a connected component of the complement of the union of the hyperplanes in \( \mathcal{A} \). Let \( r(\mathcal{A}) \) denote the number of regions for an arrangement \( \mathcal{A} \). It is natural to ask: What is the maximal number of regions for an arrangement with \( m \) hyperplanes in \( \mathbb{R}^n \)? The following Zaslavsky’s theorem answers this question.

**Proposition 1** (Zaslavsky’s Theorem (Zaslavsky, 1975; Stanley, 2004)). Let \( \mathcal{A} = \{ H_i \subset V : 1 \leq i \leq m \} \) be an arrangement in \( \mathbb{R}^n \). Then, the number of regions for the arrangement \( \mathcal{A} \) satisfies

\[
r(\mathcal{A}) \leq \sum_{i=0}^{n} \binom{m}{i}. 
\]

Furthermore, the above equality holds iff \( \mathcal{A} \) is in general position, i.e., (i) \( \dim(\bigcap_{j=1}^{k} H_{i_j}) = n - k \) for any \( k \leq n \) and \( 1 \leq i_1 < i_2 < \cdots < i_j \leq m \); (ii) \( \bigcap_{j=1}^{k} H_{i_j} = \emptyset \) for any \( k > n \) and \( 1 \leq i_1 < i_2 < \cdots < i_j \leq m \).

For example, if \( n = 2 \) then a set of lines is in general position if no two are parallel and no three meet at a point. In this case, the number of regions of an arrangement \( \mathcal{A} \) with \( m \) lines in general position is equal to

\[
r(\mathcal{A}) = \binom{m}{2} + m + 1. 
\]

For an arrangement \( \mathcal{A} \) and some \( H_0 \in \mathcal{A} \), we define

\[
\mathcal{A}^{H_0} := \{ H \cap H_0 : H \in \mathcal{A}, H \neq H_0, \ H \cap H_0 \neq \emptyset \}
\]

to be the set of nonempty intersections of \( H_0 \) and other hyperplanes in \( \mathcal{A} \). The following lemma gives a recursive method to compute \( r(\mathcal{A}) \).

**Lemma 1** (Lemma 2.1 from (Stanley, 2004)). Let \( \mathcal{A} \) be an arrangement and \( H_0 \in \mathcal{A} \). Then we have

\[
r(\mathcal{A}) = r(\mathcal{A} \setminus \{ H_0 \}) + r(\mathcal{A}^{H_0}).
\]

Lemma 1 means that we can calculate the number of regions of an arrangement by induction.

Let \# \( \mathcal{A} \) be the number of hyperplanes in \( \mathcal{A} \) and \( \text{rank}(\mathcal{A}) \) be the dimension of the space spanned by the normal vectors of the hyperplanes in \( \mathcal{A} \). An arrangement \( \mathcal{A} \) is called central if \( \bigcap_{H \in \mathcal{A}} H \neq \emptyset \).

---

1Mohamed bin Zayed University of Artificial Intelligence, UAE
2Inception Institute of Artificial Intelligence, Abu Dhabi, UAE.

Correspondence to: Huan Xiong <huan.xiong@mbzuai.ac.ae>.

Lemma 2 (Theorems 2.4 and 2.5 from (Stanley, 2004)). Let \( \mathcal{A} \) be an arrangement in an \( n \)-dimensional vector space. Then we have
\[
    r(\mathcal{A}) = \sum_{\mathcal{B} \subseteq \mathcal{A}} (-1)^{|\mathcal{B}|-\text{rank}(\mathcal{B})}.
\]

2. Proofs of Results for One-Layer CNNs

Let \([n,m] := \{n+1, n+2, \ldots, m\}\) be the set of integers from \(n\) to \(m\) and \([m] := \{1, m\} = \{1, 2, \ldots, m\}\). We establish the following generalization of Zaslavsky’s theorem, which is crucial in the proof of Theorem 2.

**Proposition 2.** Let \( V = \mathbb{R}^n, V_1, V_2, \ldots, V_m \) be \( m \) nonempty subspaces of \( V \), and \( n_1, n_2, \ldots, n_m \in \mathbb{N} \) be some nonnegative integers. Let \( \mathcal{A} = \{ H_{k,j} : 1 \leq k \leq m, \ 1 \leq j \leq n_k \} \) be an arrangement in \( \mathbb{R}^n \) with \( H_{k,j} = \{ X \in V : \alpha_{k,j} \cdot X = b_{k,j} \} \) where \( 0 \neq \alpha_{k,j} \in V_k, b_{k,j} \in \mathbb{R} \). Then, the number of regions for the arrangement \( \mathcal{A} \) satisfies
\[
    r(\mathcal{A}) \leq \sum_{(i_1, i_2, \ldots, i_m) \in K_{V;1,2,\ldots,m}} \prod_{k=1}^{m} \left( \frac{n_k}{i_k} \right),
\]
where
\[
    K_{V;1,2,\ldots,m} = \left\{(i_1, i_2, \ldots, i_m) : i_k \in \mathbb{N}, \sum_{k \in J} i_k \leq \dim \left( \sum_{k \in J} V_k \right) \forall J \subseteq [m] \right\}.
\]

Furthermore, assume that the following two conditions hold for the arrangement \( \mathcal{A} \):

(i) For each \((i_1, i_2, \ldots, i_m) \in K_{V;1,2,\ldots,m}, \) any \( \sum_{k=1}^{m} i_k \) vectors with \( i_k \) distinct vectors chosen from the set \( \{\alpha_{k,j} : 1 \leq j \leq n_k \} \) are linear independent;

(ii) For each \((i_1, i_2, \ldots, i_m) \in \mathbb{N}^m \setminus K_{V;1,2,\ldots,m}, \) the intersection of any \( \sum_{k=1}^{m} i_k \) hyperplanes with \( i_k \) distinct hyperplanes chosen from the set \( \{ H_{k,j} : 1 \leq j \leq n_k \} \) is empty.

Then, the equality in (3) holds:
\[
    r(\mathcal{A}) = \sum_{(i_1, i_2, \ldots, i_m) \in K_{V;1,2,\ldots,m}} \prod_{k=1}^{m} \left( \frac{n_k}{i_k} \right).\]

**Proof.** First, we will prove (3) by induction on \( \sum_{k=1}^{m} n_k \). When \( \sum_{k=1}^{m} n_k = 0 \), both sides of (3) equals 1 since \( \binom{0}{0} = 1 \). When \( \sum_{k=1}^{m} n_k = 1 \), both sides equals 2 since \( \binom{1}{0} + \binom{1}{1} = 2 \). Suppose that the result is true for \( \sum_{k=1}^{m} n_k \leq N \) for some \( N \geq 1 \). Now consider the case \( \sum_{k=1}^{m} n_k = N + 1 \). Without loss of generality, assume \( n_1 \geq 1 \). Then \( H_{1,1} \in \mathcal{A} \). Notice that the translation \( Y \rightarrow Y + Y_0 \) for some \( Y_0 \in \mathbb{R}^n \) (i.e., translate all points in \( \mathbb{R} \) by a vector \( Y_0 \)) doesn’t change the number of regions in \( \mathcal{A} \). Thus we can assume \( b_{1,1} = 0 \). Then \( H_{1,1} \) becomes an \( (n-1) \)-dimensional subspace of \( V \). Replace \( H_0 \) in Lemma 1 with \( H_{1,1} \), we obtain
\[
    r(\mathcal{A}) = r(\mathcal{A} \setminus \{H_{1,1}\}) + r(\mathcal{A}_{H_{1,1}}) .
\]

By induction hypothesis, we have
\[
    r(\mathcal{A} \setminus \{H_{1,1}\}) \leq \sum_{(i_1, i_2, \ldots, i_m) \in K_{V;1,2,\ldots,m}} \left( \frac{n_1-1}{i_1} \right) \prod_{k=2}^{m} \left( \frac{n_k}{i_k} \right),
\]
and
\[
    r(\mathcal{A}_{H_{1,1}}) \leq \sum_{(i_1, i_2, \ldots, i_m) \in K_{V \cap H_{1,1};1,2,\ldots,m}} \left( \frac{n_1-1}{i_1} \right) \prod_{k=2}^{m} \left( \frac{n_k}{i_k} \right) .
\]

Let’s consider (7) first. Since \( H_{1,1} \) is the orthogonal complement of the linear subspace generated by \( \alpha_{1,1} \), and \( 0 \neq \alpha_{1,1} \subset V_1 \), we have
\[
    H_{1,1} + V_1 = V.
\]
Let $V'_k = H_{1,1} \cap V_k$ for $1 \leq k \leq m$. Therefore, for each $J \subseteq [2, m]$, we have

$$\dim \left( H_{1,1} \cap \left( V_1 + \sum_{k \in J} V_k \right) \right) = \dim(H_{1,1}) + \dim \left( V_1 + \sum_{k \in J} V_k \right) - \dim(V) = \dim \left( V_1 + \sum_{k \in J} V_k \right) - 1 \quad (8)$$

and thus

$$\dim \left( V'_1 + \sum_{k \in J} V'_k \right) = \dim \left( V_1 + \sum_{k \in J} V_k \right) - 1. \quad (9)$$

On the other hand, it is trivial that

$$\dim \left( \sum_{k \in J} V'_k \right) \leq \dim \left( \sum_{k \in J} V_k \right) \quad (10)$$

for any $J \subseteq [2, m]$. Therefore, by (7) we derive

$$r(A^{H_{1,1}}) \leq \sum_{(i_1, i_2, \ldots, i_m) \in K_{V_{1,1}, V'_1, \ldots, V'_m}} \left( \begin{array}{c} n_1 - 1 \\ i_1 \\ \end{array} \right) \prod_{k=2}^{m} \left( \begin{array}{c} n_k \\ i_k \\ \end{array} \right)$$

$$\leq \sum_{i_1-1 + \sum_{k \in J} i_k \leq \dim \left( V_1 + \sum_{k \in J} V_k \right) \forall J \subseteq [2, m]} \left( \begin{array}{c} n_1 - 1 \\ i_1 - 1 \\ \end{array} \right) \prod_{k=2}^{m} \left( \begin{array}{c} n_k \\ i_k \\ \end{array} \right)$$

$$\leq \sum_{i_1 + \sum_{k \in J} i_k \leq \dim \left( V_1 + \sum_{k \in J} V_k \right) \forall J \subseteq [2, m]} \left( \begin{array}{c} n_1 - 1 \\ i_1 - 1 \\ \end{array} \right) \prod_{k=2}^{m} \left( \begin{array}{c} n_k \\ i_k \\ \end{array} \right)$$

$$= \sum_{(i_1, i_2, \ldots, i_m) \in K_{V_1, V_2, \ldots, V_m}} \left( \begin{array}{c} n_1 - 1 \\ i_1 - 1 \\ \end{array} \right) \prod_{k=2}^{m} \left( \begin{array}{c} n_k \\ i_k \\ \end{array} \right). \quad (11)$$

Put (5), (6) and (11) together, we obtain

$$r(A) \leq \sum_{(i_1, i_2, \ldots, i_m) \in K_{V_1, V_2, \ldots, V_m}} \left( \prod_{k=1}^{m} \left( \begin{array}{c} n_k \\ i_k \\ \end{array} \right) \right)$$

$$= \sum_{(i_1, i_2, \ldots, i_m) \in K_{V_1, V_2, \ldots, V_m}} \prod_{k=1}^{m} \left( \begin{array}{c} n_k \\ i_k \\ \end{array} \right), \quad (12)$$

which competes the proof of (3).

Furthermore, assume that the arrangement $A$ satisfies the condition (i) and (ii). Then, the central sub-arrangements $B$ consisting of $\sum_{k=1}^{m} i_k$ hyperplanes with $i_k$ distinct hyperplanes chosen from the set $\{H_{k,j} : 1 \leq j \leq n_k\}$, where $(i_1, i_2, \ldots, i_m) \in K_{V_{1,1}, V_2, \ldots, V_m}$. In this case, $\#B = \text{rank}(B) = \sum_{k=1}^{m} i_k$. Also, for any given $(i_1, i_2, \ldots, i_m) \in K_{V_1, V_2, \ldots, V_m}$, we have $\left( \begin{array}{c} n_k \\ i_k \end{array} \right)$ choices to pick $i_k$ hyperplanes from each $\{\alpha_{k,i} : 1 \leq i \leq n_k\}$. Therefore, by Lemma 2 we obtain

$$r(A) = \sum_{B \subseteq A, B \text{ central}} (-1)^{\#B - \text{rank}(B)} = \sum_{B \subseteq A, B \text{ central}} 1 = \sum_{(i_1, i_2, \ldots, i_m) \in K_{V_1, V_2, \ldots, V_m}} \prod_{k=1}^{m} \left( \begin{array}{c} n_k \\ i_k \end{array} \right).$$
To prove Theorem 2, we need the following lemmas on picking distinct elements from the union of certain sets.

**Lemma 3.** Let \( S_1, S_2, \ldots, S_m \) be \( m \) finite sets, and \( a_1, a_2, \ldots, a_m \) be some nonnegative integers such that for any \( I \subseteq [m] \),
\[
\sum_{i \in I} a_i \leq \# \bigcup_{i \in I} S_i.
\]
(13)

Then, we can take \( a_i \) elements from each \( S_i \) such that these \( \sum_{i=1}^m a_i \) elements are distinct.

**Proof.** We will prove this lemma by induction on \( m \). When \( m = 1 \), the claim is trivial. Now assume that the lemma holds for any \( 1 \leq m < n \) and consider the case \( m = n \). Without loss of generality, we assume that there exists some \( \emptyset \neq I \subseteq [n] \) such that (otherwise we can always increase some \( a_i \) to make the following equality holds for some \( I \))
\[
\sum_{i \in I} a_i = \# \bigcup_{i \in I} S_i.
\]
(14)

The proof is divided into two cases.

Case (1): There exists some \( I \) satisfying (14) with \( \emptyset \neq I \neq [n] \). In this case, we can assume that \( I = [r] \) for some \( 1 \leq r \leq n-1 \) by symmetry, i.e.,
\[
\sum_{i=1}^r a_i = \# \bigcup_{i=1}^r S_i.
\]
(15)

Let
\[
S'_j = S_{j+r} \setminus \bigcup_{i=1}^r S_i, \quad 1 \leq j \leq n-r.
\]

Then \( \bigcup_{j \in J} S'_j \cap \bigcup_{i=1}^r S_i = \emptyset \). Therefore, for any \( J \subseteq [n-r] \), we have
\[
\# \bigcup_{j \in J} S'_j = \# \left( \bigcup_{j \in J} S'_j \cup \bigcup_{i=1}^r S_i \right) - \# \bigcup_{i=1}^r S_i = \# \left( \bigcup_{j \in J} S_{j+r} \cup \bigcup_{i=1}^r S_i \right) - \# \bigcup_{i=1}^r S_i.
\]
(16)

By (13) and (15) the above equality becomes
\[
\# \bigcup_{j \in J} S'_j \geq \left( \sum_{j \in J} a_{j+r} + \sum_{i=1}^r a_i \right) - \sum_{i=1}^r a_i = \sum_{j \in J} a_{r+j}.
\]
(17)

Since \( 1 \leq \# I \leq n-1 \), by induction we can pick \( a_i \) elements from each \( S_i \) for \( 1 \leq i \leq r \), and \( a_{r+j} \) elements from each \( S_{j+r} \) for \( 1 \leq j \leq n-r \) such that these \( \sum_{i=1}^n a_i \) elements are distinct. Thus the claim holds.

Case (2): The only \( I \) satisfying (14) is \( I = [n] \). Then \( \# S_1 > a_1 \) and thus \( S_1 \cap \bigcup_{i=2}^n S_i \neq \emptyset \) (otherwise \( \sum_{i=1}^n a_i = \# \bigcup_{i=1}^n S_i = \# S_1 + \# \bigcup_{i=2}^n S_i > \sum_{i=1}^n a_i \), a contradiction). Let \( x \in S_1 \cap \bigcup_{i=2}^n S_i \) and
\[
S'_j = \begin{cases} S_j, & 2 \leq j \leq n; \\ S_j \setminus \{x\}, & j = 1. \end{cases}
\]

Then \( \{S'_j : 1 \leq j \leq n\} \) still satisfies (13). But \( \sum_{i=1}^n \# S'_j < \sum_{i=1}^n \# S_i \). Then \( \{S'_j : 1 \leq j \leq n\} \) either satisfies Case (1), which leads to a solution; or still in Case (2), which we can continue the process until Case (i) satisfies. This completes the proof.

**Lemma 4.** Let \( S_1, S_2, \ldots, S_m \) be \( m \) finite sets. Then, there exist some \( a_1, a_2, \ldots, a_m \in \mathbb{N} \) such that
\[
\sum_{i=1}^m a_i = \# \bigcup_{i=1}^m S_i.
\]
(18)

and for any \( I \subseteq [m] \),
\[
\sum_{i \in I} a_i \leq \# \bigcup_{i \in I} S_i.
\]
(19)
Supplementary Material

Proof. We will prove it by Induction on $m$. The claim is trivial when $m = 1$. Now assume that $m \geq 2$ and the result is true for $m - 1$. Therefore, we can pick some $a_1, a_2, \ldots, a_{m-1} \in \mathbb{N}$ such that

$$
\sum_{i=1}^{m-1} a_i = \# \bigcup_{i=1}^{m-1} S_i, \tag{20}
$$

and for any $I \subseteq [m - 1],$

$$
\sum_{i \in I} a_i \leq \# \bigcup_{i \in I} S_i. \tag{21}
$$

Furthermore, let $a_m = \# \left( S_m \setminus \bigcup_{i=1}^{m-1} S_i \right)$. Then, for any $I \subseteq [m - 1]$, we have

$$
a_m + \sum_{i \in I} a_i \leq \# \bigcup_{i \in I} S_i + \# \left( S_m \setminus \bigcup_{i=1}^{m-1} S_i \right) \leq \# \bigcup_{i \in I \cup \{m\}} S_i. \tag{22}
$$

Also,

$$
\sum_{i=1}^{m} a_i = \# \bigcup_{i=1}^{m-1} S_i + \# \left( S_m \setminus \bigcup_{i=1}^{m-1} S_i \right) = \# \bigcup_{i=1}^{m} S_i. \tag{23}
$$

Then the claim is also true for $m$. \hfill \square

We also need the following lemmas on measure zero subsets of Euclidean spaces with respect to Lebesgue measure.

Lemma 5. Let $V \cong \mathbb{R}^n$ be a vector space. Then $S = \{(v_1, v_2, \ldots, v_n) \in V^n : v_1, v_2, \ldots, v_n \text{ are linear dependent}\}$ is a measure zero subset of $V^n$, with respect to Lebesgue measure.

Proof. Without loss of generality, assume $V = \mathbb{R}^n$. Let the $i$-th vector be $v_i = (x_{i,1}, x_{i,2}, \ldots, x_{i,n})$. Then $v_1, v_2, \ldots, v_n$ are linear dependent iff $det((x_{i,j})_{n \times n}) = 0$,

whose left hand side is a non-zero polynomial of all $x_{i,j}$. It is easy to see that the solution of this polynomial has co-dimension 1 in $\mathbb{R}^{n \times n}$, thus $S$ is a measure zero set. \hfill \square

Lemma 6. Let $m > n$ be two given positive integers, $A = (a_{ij})_{m \times n} \in \mathbb{R}^{m \times n}$ and $C = (c_1, c_2, \ldots, c_m) \in \mathbb{R}^m$. Let $S$ be the set of $(A, C) \in \mathbb{R}^{m(n+1)}$ such that

$$
\begin{align*}
& a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = c_1 \\
& a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = c_2 \\
& \vdots \\
& a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = c_m
\end{align*}
$$

has solutions for $(x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$. Then $S$ is a measure zero subset of $\mathbb{R}^{m(n+1)}$, with respect to Lebesgue measure.

Proof. By Lemma 5, the augmented matrix $(A, C)$ has the rank $(n + 1)$ except for a measure zero subset of $\mathbb{R}^{m(n+1)}$. On the other hand, the rank of the matrix $A$ is at most $n$. Therefore, the rank of the augmented matrix $(A, C)$ is larger than the rank of $A$ except for a measure zero subset of $\mathbb{R}^{m(n+1)}$, thus by Rouché-Capelli Theorem (Shafarevich & Remizov, 2012) we obtain that (6) has no solutions except for a measure zero set of $\mathbb{R}^{m(n+1)}$. \hfill \square

Lemma 3 implies the following results when we choose a basis of a linear space properly.

[Insert additional content here if necessary]
Lemma 7. Let $V \cong \mathbb{R}^n$ be a vector space and $V_i$ ($1 \leq i \leq m$) be $m$ subspaces of $V$. Suppose that some non-negative integers $a_i$ ($1 \leq i \leq m$) satisfy

$$\sum_{i \in I} a_i \leq \dim(\sum_{i \in I} V_i)$$

for each $I \subseteq [m]$. Then we obtain the following result.

(i) We can pick $a_i$ vectors from $V_i$ for $1 \leq i \leq m$ such that these $\sum_{1 \leq i \leq m} a_i$ vectors are linear independent.

(ii) $\sum_{1 \leq i \leq m} a_i$ vectors with $a_i$ vectors from $V_i$ for $1 \leq i \leq m$ such that they are linear dependent, forms a measure zero set in $\prod_{i=1}^m V_i^{a_i}$, with respect to Lebesgue measure.

Proof. (i) By linear algebra, we can construct a basis $v_1, v_2, \ldots, v_n$ of $V$ such that each $V_i$ has a basis which is a subset of $v_1, v_2, \ldots, v_n$. Then, by Lemma 3 this claim holds.

(ii) Let $n' = \sum_{1 \leq i \leq m} a_i$. By (i) there exist $n'$ linear independent vectors $v_1, v_2, \ldots, v_{n'}$ with $a_i$ vectors from $V_i$ for $1 \leq i \leq m$. Let $V'_i$ be the vector spaces generated by such $a_i$ vectors in $V_i$. For any $n'$ linear dependent vectors $v'_1, v'_2, \ldots, v'_{n'}$ with $a_i$ vectors from $V_i$ for $1 \leq i \leq m$, their projections $v''_1, v''_2, \ldots, v''_{n'}$ onto $\prod_{i=1}^m V'_i$ are also linear dependent. Suppose that $v''_k = \sum_{j=1}^{n'} y_{k,j} v_j$ for $1 \leq k \leq n'$. If $v''_k$ are chosen from $V_i$, such that $v_j \notin V_i$, we set $y_{k,j} = 0$. Otherwise, we set $y_{k,j} = y_{j,k}$. Therefore, $\# \{y_{k,j}\}$ equals the dimension of the projection of $\prod_{i=1}^m V_i^{a_i}$ onto $\prod_{i=1}^m V'_i$. Also, $v'_1, v'_2, \ldots, v'_{n'}$ are linear dependent iff

$$\det ((y_{k,j})_{n' \times n'}) = 0.$$

Since $v_1, v_2, \ldots, v_{n'}$ are linear independent, the left hand side $\det ((y_{k,j})_{n' \times n'})$ must be a non-zero polynomial of some $y_{k,j}$. Therefore, the solution of this polynomial forms a measure zero set in $\mathbb{R}^n \setminus \{y_{k,j}\}$ due to the zero measurability of the solutions of non-zero polynomial in Euclidean spaces (see (Lojasiewicz, 1964)). Thus such $\sum_{1 \leq i \leq m} a_i$ vectors forms a measure zero set in $\prod_{i=1}^m V_i^{a_i}$, with respect to Lebesgue measure. $\blacksquare$

Now we are ready to prove Theorem 2.

Proof of Theorem 2. By Definition 1, the number of linear regions of $\mathcal{N}$ at $\theta$ is equal to the number of regions of the hyperplane arrangement

$$\mathcal{A}_{\mathcal{N}, \theta} := \{ H_{i,j,k}(X^0; \theta) : 1 \leq i \leq n_1^{(1)}, 1 \leq j \leq n_1^{(2)}, 1 \leq k \leq d_1 \},$$

where $H_{i,j,k}(X^0; \theta)$ is the hyperplane determined by $Z_{i,j,k}^1(X^0; \theta) = 0$ (the expression of $Z_{i,j,k}^1(X^0; \theta)$ is given in (2)). Recall that $X^0 = (X_{a,b,c}^0)_{a_0^{(1)} \times n_0^{(2)} \times d_0}$. Then $H_{i,j,k}(X^0; \theta)$ can be written as

$$\langle \alpha_{i,j,k}, X^0 \rangle_F + B^{1,k} = 0,$$

where $\langle \cdot, \cdot \rangle_F$ is the Frobenius inner product, $\alpha_{i,j,k}$ is an $n_0^{(1)} \times n_0^{(2)} \times d_0$ dimensional tensor, whose $(a + (i-1)s_1, b + (j-1)s_1, c)$-th element is $W_{a,b,c}^{1,k}$ for all $1 \leq a \leq f^1_1$, $1 \leq b \leq f^2_1$, $1 \leq c \leq d_0$; and 0 otherwise. Let

$$V_{i,j} = \{ \beta \in \mathbb{R}^{n_0^{(1)} \times n_0^{(2)} \times d_0} : \beta_{a',b',c'} = 0 \ \forall (a', b', c') \neq (a + (i-1)s_1, b + (j-1)s_1, c) \}$$

be the subspace of $\mathbb{R}^{n_0^{(1)} \times n_0^{(2)} \times d_0}$ generated by $n_0^{(1)} \times n_0^{(2)} \times d_0$ dimensional tensors whose $(a + (i-1)s_1, b + (j-1)s_1, c)$-th element ranges over $\mathbb{R}$ for all $1 \leq a \leq f^1_1$, $1 \leq b \leq f^2_1$, $1 \leq c \leq d_0$; and 0 otherwise. Then $\alpha_{i,j,k} \in V_{i,j}$ for $1 \leq k \leq d_1$.

By Proposition 2, we obtain

$$R_{\mathcal{N}, \theta} = r(\mathcal{A}_{\mathcal{N}, \theta}) \leq \sum_{(i,j) \in I_{\mathcal{N}} : (i,j) \in K_{\mathcal{N}}(V_{i,j})} \prod_{k=1}^m \binom{d_1}{d_1},$$

(24)
where
\[
K_{V}(\{(i,j)\in I_{N}: \sum_{(i,j)\in J} t_{i,j} \leq \dim \left( \sum_{(i,j)\in J} V_{i,j} \right) \forall J \subseteq I_{N} \}) = \{(t_{i,j})_{(i,j)\in I_{N}}: \sum_{(i,j)\in J} t_{i,j} \neq \# \cup_{(i,j)\in J} S_{i,j} \forall J \subseteq I_{N} \},
\]

which gives an upper bound for $R_{N,\theta}$ and $R_{N'}$. Next we will show that this upper bound can be reached except for a measure zero set in $\mathbb{R}^{\text{#weights}+\text{#bias}}$ with respect to Lebesgue measure. By Lemmas 6 and 7, when $\theta$ ranges over $\mathbb{R}^{\text{#weights}+\text{#bias}}$, the set of $\theta$ such that $A_{N',\theta}$ satisfies the conditions (i) and (ii) of Proposition 2 (replace $\{i_{k}: 1 \leq k \leq m\}$ by $\{(i_{j},j)\in I_{N}\}$, and $\{V_{k}: 1 \leq k \leq m\}$ by $\{V_{i,j}: (i,j)\in I_{N}\}$), forms a complement of a measure zero set in $\mathbb{R}^{\text{#weights}+\text{#bias}}$, with respect to Lebesgue measure. Then, for such parameters $\theta$, by Proposition 2 we derive the equality holds for (24), which implies that the maximal number $R_{N'}$ of linear regions of $N'$ is equal to
\[
R_{N'} = \sum_{(i,j)\in I_{N}} \prod_{(i,j)\in I_{N}} \left( d_{1} - t_{i,j} \right),
\]

and the right hand side of the above equality also equals the expectation of the number $R_{N',\theta}$ of linear regions of $N'$ with respect to the distribution $\mu$ of weights and biases. \hfill \square

The following result gives a simple example for Theorem 2.

**Corollary 1.** Let $N$ be a one-layer ReLU CNN with input dimension $1 \times n \times 1$. Assume there are $d_{1}$ filters with dimension $1 \times 2 \times 1$ and stride $s = 1$. Thus the hidden layer dimension is $1 \times (n-1) \times d_{1}$. When $n$ is fixed, we have
\[
R_{N} = \frac{(n-1)}{2} d_{1}^{n} + O(d_{1}^{n-1}).
\]

**Proof.** By Theorem 2, we obtain
\[
R_{N} = \sum_{(i,j)\in I_{N}} \prod_{(i,j)\in I_{N}} \left( d_{1} - t_{i,j} \right).
\]

Furthermore, when $n$ is fixed, $R_{N}$ is a polynomial of $d_{1}$ with degree $n$ by Lemma 3 in the main paper. To calculate the coefficient of the leading term $d_{1}^{n}$ of this polynomial, we need to determine all $(t_{i,j})_{(i,j)\in I_{N}} \in K_{N}$ with $\sum_{(i,j)\in I_{N}} t_{i,j} = n$. First, since $n_{1}^{(1)} = 1$ and $n_{1}^{(2)} = n-1$, it is easy to see that $I_{N} = \{(1,j): 1 \leq j \leq n-1\}$ and $S_{1,j} = \{(1,j,1), (1,j+1,1)\}$ for each $1 \leq j \leq n-1$. Therefore,
\[
K_{N} = \{(t_{1,j})_{1\leq j\leq n-1}: t_{1,j} \in \mathbb{N}, \sum_{j\in J} t_{1,j} \leq \# \cup_{(1,j)\in J} S_{1,j} \forall J \subseteq [n-1]\}.
\]

Then, there are $n-1$ vectors $(t_{1,j})_{1\leq j\leq n-1} \in K_{N}$ satisfying $\sum_{j=1}^{n-1} t_{1,j} = n$: $(2,1,1,\ldots,1)$, $(1,2,1,\ldots,1)$, $(1,1,2,1,\ldots,1)$, $\ldots$, $(1,1,1,\ldots,1,2)$. Therefore, the leading term in $R_{N}$ equals
\[
(n-1)\left( \frac{d_{1}}{2} \right) d_{1}^{n-2} = \frac{(n-1)}{2} d_{1}^{n} - \frac{(n-1)}{2} d_{1}^{n-1}
\]

and thus
\[
R_{N} = \frac{(n-1)}{2} d_{1}^{n} + O(d_{1}^{n-1}).
\]

This completes the proof. \hfill \square

Next, we prove Lemma 3 and Theorem 3 in the main paper.
Supplementary Material

Proof of Lemma 3 in the main paper. Directly replace \( \{a_i : 1 \leq i \leq m\} \) by \( \{t_{i,j} : (i,j) \in I_N\} \), and \( \{S_i : 1 \leq i \leq m\} \) by \( \{S_{i,j} : (i,j) \in I_N\} \) in Lemma 4, we derive the result.

Proof of Theorem 3. It is easy to see that \( (d^1_{t_{i,j}}) = \Theta(d^1_{t_{i,j}}) \) when \( d_1 \) tends to infinity. Then, by Eq. (4) and Lemma 3 in the main paper, we have

\[
R_N = \Theta(d^1_{t_{i,j}}). 
\]  

Furthermore, if all input neurons have been involved in the convolutional calculation, we have

\[
\cup_{(i,j) \in I_N} S_{i,j} = \{(a,b,c) : 1 \leq a \leq n_0^{(1)}, 1 \leq b \leq n_0^{(2)}, 1 \leq c \leq d_0\} \tag{30}
\]

and thus

\[
R_N = \Theta(d^1_{T_{0}^{(1)}}, n_0^{(2)} \times d_0). 
\]

3. Proofs of Results for Multi-Layer CNNs

In this section, we prove Theorem 5 on multi-layer ReLU CNNs.

Proof of Theorem 4. Assume that the parameters \( W \) and \( B \) for such two convolutional layers are the same as defined in Section 2. Let \( l = 1, 2 \) in (2) in the main paper and \( X^l_{i,j,k} = Z_{i,j,k}^l(X^0; \theta) \), we obtain

\[
X^1_{i,j,k} = \sum_{a=1}^{f^{(1)}_{1}} \sum_{b=1}^{f^{(2)}_{1}} \sum_{c=1}^{d_0} W^{1,k}_{a,b,c} X^0_{a+(i-1)s_1,b+(j-1)s_1,c} + B^{1,k} \tag{31}
\]

and

\[
X^2_{i,j,k} = \sum_{a=1}^{f^{(1)}_{2}} \sum_{b=1}^{f^{(2)}_{2}} \sum_{c=1}^{d_0} W^{2,k}_{a,b,c} X^0_{a+(i-1)s_2,b+(j-1)s_2,c} + B^{2,k} \tag{32}
\]

Substitute (31) into (32), we derive

\[
X^2_{i,j,k} = \sum_{a'=1}^{f^{(1)}_{2}} \sum_{b'=1}^{f^{(2)}_{2}} \sum_{c'=1}^{d_0} \sum_{a=1}^{f^{(1)}_{1}} \sum_{b=1}^{f^{(2)}_{1}} \sum_{c=1}^{d_0} W^{2,k}_{a',b',c'} W^{1,c'}_{a,b,c} X^0_{a+(i-1)s_2,b+(j-1)s_2-1,s_1,c} + \text{const} \tag{33}
\]

\[
= \sum_{a'=1}^{f^{(1)}_{2}} \sum_{b'=1}^{f^{(2)}_{2}} \sum_{c'=1}^{d_0} \sum_{a=1}^{f^{(1)}_{1}} \sum_{b=1}^{f^{(2)}_{1}} \sum_{c=1}^{d_0} W^{2,k}_{a',b',c'} W^{1,c'}_{a,b,c} X^0_{a+(i-1)s_2,b+(j-1)s_2,s_1 + \text{const}.} \tag{34}
\]

Note that \( 1 \leq a + (a'-1)s_1 \leq f^{(1)}_{1} + (f^{(2)}_{1} - 1)s_1 \) and \( 1 \leq b + (b'-1)s_1 \leq f^{(2)}_{1} + (f^{(2)}_{2} - 1)s_1 \). Then (33) becomes

\[
X^2_{i,j,k} = \sum_{a=1}^{f^{(1)}_{1} + (f^{(2)}_{1} - 1)s_1} \sum_{b=1}^{f^{(2)}_{1} + (f^{(2)}_{2} - 1)s_1} \sum_{c=1}^{d_0} W^{rk}_{a,b,c} X^0_{a+(i-1)s_2,b+(j-1)s_2,c} + \text{const} \tag{35}
\]

where \( W^{rk}_{a,b,c} \) are some constants. Therefore, \( N \) is realized as a ReLU CNN with one hidden convolutional layer such that its \( d_2 \) filters has size \( (f^{(1)}_{1} + (f^{(2)}_{1} - 1)s_1) \times (f^{(2)}_{1} + (f^{(2)}_{2} - 1)s_1) \times d_0 \) and stride \( s_1s_2 \), which completes the proof.

Proof of Theorem 5. (i) The basic idea is to map many regions of the input space of each layer to the same set, thus identify many regions of space.
When \( \phi \) which means that

\[
W_{a,b,c}^{1,k} = \begin{cases} 
1, & \text{if } a = b = 1, k = (c-1)p + 1, \ 1 \leq c \leq d_0; \\
2, & \text{if } a = b = 1, (c-1)p + 2 \leq k \leq cp, 1 \leq c \leq d_0; \\
0, & \text{otherwise}
\end{cases}
\]  

and

\[
B^{1,k} = \begin{cases} 
-(k - (c-1)p - 1), & \text{if } (c-1)p + 1 \leq k \leq cp \text{ for some } 1 \leq c \leq d_0; \\
0, & \text{otherwise.}
\end{cases}
\]

Therefore, by (2) in the main paper we obtain

\[
Z_{i,j,k}^1(X^0; \theta) = \begin{cases} 
X_{1+i-(i-1)s_1,1+(j-1)s_1,c}^0, & \text{if } k = (c-1)p + 1 \text{ for some } 1 \leq c \leq d_0; \\
2X_{1+i-(i-1)s_1,1+(j-1)s_1,c}^0 - (k - (c-1)p - 1), & \text{if } (c-1)p + 2 \leq k \leq cp \text{ for some } 1 \leq c \leq d_0; \\
0, & \text{otherwise.}
\end{cases}
\]  

When \(W_{a,b,c}^{1,k}\) and \(B^{1,k}\) are given as in (36) and (37), the map

\[
X_{i,j,k}^1 = \max\{0, Z_{i,j,k}^1(X^0; \theta)\}
\]
determines a function

\[
X^1 = \Phi_1(X^0)
\]

from \(\mathbb{R}^{n_0(1)} \times n_0(2) \times d_0\) to \(\mathbb{R}^{n_1(1)} \times n_1(2) \times d_1\).

For each \(i, j \in \mathbb{N}^+\), let

\[
\psi_i(x) = \begin{cases} 
\max\{0, x\}, & \text{if } i = 1; \\
\max\{0, 2x - (i-1)\}, & \text{if } i \geq 2
\end{cases}
\]

and

\[
\phi_j(x) = \sum_{i=1}^j (-1)^{i+1}\psi_i(jx).
\]

Then it is easy to check that

\[
\phi_j(x) = \begin{cases} 
0, & \text{if } x \leq 0; \\
 jx - i, & \text{if } \frac{j}{j} \leq x \leq \frac{2i+1}{2j} \leq \frac{i}{j} \text{ where } i \in \mathbb{N}; \\
 i - jx, & \text{if } \frac{2i-1}{2j} \leq x \leq \frac{i}{j} \leq \frac{i}{j} \text{ where } i \in \mathbb{N}^+,
\end{cases}
\]

which means that \(\phi_j\) is an affine function when restricted to each interval \([0, \frac{1}{2j}], [\frac{1}{2j}, \frac{2}{2j}], \ldots, [\frac{j-1}{2j}, \frac{j}{2j}]\) and furthermore \(\phi_j([0, \frac{1}{2j}]) = \phi_j([\frac{1}{2j}, \frac{2}{2j}]) = \cdots = \phi_j([\frac{j-1}{2j}, \frac{j}{2j}]) = [0, \frac{1}{2}]\) (i.e., \(\phi_j(x)\) sends \(j\) distinct intervals \([0, \frac{1}{2j}], [\frac{1}{2j}, \frac{2}{2j}], \ldots, [\frac{j-1}{2j}, \frac{j}{2j}]\) to the same interval \([0, \frac{1}{2}]\).

Next, we define an intermediate convolutional layer (without activation functions) from

\[
X^1 = (X_{a,b,c}^1)_{n_1^{(1)} \times n_1^{(2)} \times d_1}
\]
to

\[ Y^1 = (Y^1_{a,b,c})_{n_1^{(1)} \times n_1^{(2)} \times d_0} \]

between the first and second hidden convolutional layers. We set the \( d_0 \) filters with size \( 1 \times 1 \times d_1 \), the stride 1, and define the weights \( W' \) and biases \( B' \) in this intermediate convolutional layer as

\[
W'^{1,c}_{1,1,k} = \begin{cases} 
p \cdot (-1)^{i+1}, & \text{if } k = (c-1)p + i, \ 1 \leq c \leq d_0; \\
0, & \text{otherwise} \end{cases}
\]

and

\[
B'^{1,k} = 0 \ \forall \ 1 \leq k \leq d_0.
\]

Then by (2) in the main paper,

\[
Y^1_{a,b,c} = p \sum_{i=1}^{p} (-1)^{i+1} X^1_{a,b,(c-1)p+i}
\]

for \( 1 \leq a \leq n_1^{(1)}, \ 1 \leq b \leq n_1^{(2)}, \ 1 \leq c \leq d_0 \). Therefore, (46) determines an affine function

\[
Y^1 = \Phi'_1(X^1)
\]

from \( \mathbb{R}^{n_1^{(1)} \times n_1^{(2)} \times d_1} \) to \( \mathbb{R}^{n_1^{(1)} \times n_1^{(2)} \times d_0} \). Therefore, we obtain

\[
Y^1_{a,b,c} = p \sum_{i=1}^{p} (-1)^{i+1} X^1_{a,b,(c-1)p+i}
\]

\[
= p \sum_{i=1}^{p} (-1)^{i+1} \max \{0, Z^1_{a,b,(c-1)p+i}\}
\]

\[
= p \sum_{i=1}^{p} (-1)^{i+1} \psi_i(pX^0_{1+(a-1)s_1,1+(b-1)s_1,c})
\]

\[
= \phi_p(X^0_{1+(a-1)s_1,1+(b-1)s_1,c}).
\]

The third equality holds due to Eqs. (38) and (41). By the previous discussion on properties of the function \( \phi_j(x) \), the following map \( \Psi_1 = \Phi'_1 \circ \Phi_1 \) determined by Eq. (48)

\[ \Psi_1: \mathbb{R}^{n_0^{(1)} \times n_0^{(2)} \times d_0} \xrightarrow{\phi_{t_1}} \mathbb{R}^{n_1^{(1)} \times n_1^{(2)} \times d_1} \xrightarrow{\Phi'_1} \mathbb{R}^{n_1^{(1)} \times n_1^{(2)} \times d_0} \]

\[ X^0 \mapsto X^1 \mapsto Y^1 \]

sends \( \frac{d_1}{d_0} \) distinct hypercubes

\[ \left\{ \left[ 0, \frac{1}{2p} \right], \left[ \frac{1}{2p}, \frac{2}{2p} \right], \ldots, \left[ \frac{p-1}{2p}, \frac{p}{2p} \right] \right\}^{n_0^{(1)} \times n_0^{(2)} \times d_0} \]

in \( [0, \frac{1}{2}]^{n_1^{(1)} \times n_1^{(2)} \times d_0} \) onto the same hypercube \( [0, \frac{1}{2}]^{n_1^{(1)} \times n_1^{(2)} \times d_0} \) of the intermediate layer \( Y^1 \in \mathbb{R}^{n_1^{(1)} \times n_1^{(2)} \times d_0} \) (this map is affine and bijective when restricted to each of the \( \frac{d_1}{d_0} \) distinct hypercubes). Similarly (keep \( d_0 \) unchanged, and replace \( n_0^{(1)}, n_0^{(2)}, n_1^{(1)}, n_1^{(2)}, d_1 \) in \( \Psi_1 \) by \( n_{l-1}^{(1)}, n_{l-1}^{(2)}, n_l^{(1)}, n_l^{(2)}, d_l \)), we can define \( \Phi_l, \Phi'_l, \Psi_l \) and \( Y^l \) for \( 2 \leq l \leq L - 1 \) such that the map

\[ \Psi_l: \mathbb{R}^{n_{l-1}^{(1)} \times n_{l-1}^{(2)} \times d_{l-1}} \xrightarrow{\phi_{t_l}} \mathbb{R}^{n_l^{(1)} \times n_l^{(2)} \times d_l} \xrightarrow{\Phi'_l} \mathbb{R}^{n_l^{(1)} \times n_l^{(2)} \times d_0} \]

\[ Y^{l-1} \mapsto X^l \mapsto Y^l \]
which is calculated according to Theorem 2 which completes the proof.

(ii) We will prove this claim by induction on $l$. Let $N_{L-1}$, $N_0$ denote the intermediate layer. Note that $\Phi_l \circ \Phi_{l-1}$ is the convolutional layer between $X^{l-1}$ and $X^l$ which has $d_l$ filter with size $f^{(1)}_l \times f^{(2)}_l \times d_{l-1}$ and stride $s_l$ due to Theorem 4. Finally, by Theorem 2, a one-layer ReLU CNN with input dimension $n^{(1)}_{L-1} \times n^{(2)}_{L-1} \times d_{L-1}$ and output dimension $n^{(1)}_L \times n^{(2)}_L \times d_L$ can divide the hypercube $[0, \frac{1}{2}]^{n^{(1)}_{L-1} \times n^{(2)}_{L-1} \times d_{L-1}}$ into $R_{\mathcal{N}'}$ regions.

Put the network from $X^0$ to $Y^{L-1}$ and $Y^{L-1}$ to $X^L$ together, we prove the lower bound claim.

(4) We will prove this claim by induction on $L$. When $L = 1$, by Theorem 2 the claim is true. Now suppose that $L \geq 2$ and the claim is true for $L - 1$. Let $\mathcal{N}'$ be the CNN obtained from $\mathcal{N}$ by deleting the $L$-th hidden layer (i.e., $\mathcal{N}'$ consists of the first to the $L - 1$-th layer of $\mathcal{N}$). Then by induction hypothesis, we have

$$R_{\mathcal{N}'} \leq R_{\mathcal{N}''} \prod_{i=2}^{L-1} \sum_{i=0}^{n^{(1)}_{i-1}n^{(2)}_i \times d_i} \binom{n^{(1)}_i n^{(2)}_i}{d_i}.$$  

Now we consider the $L$-th layer. Suppose that the CNN $\mathcal{N}$ with parameters $\theta$ partitions the input space into $m$ distinct linear regions $R_i$ ($1 \leq i \leq m$). Since each linear region $R_i$ corresponds to a certain activation pattern, the function $F_{\mathcal{N}', \theta}$ becomes an affine function when restricted to $R_i$. Therefore, after adding the $L$-th layer to $\mathcal{N}'$, when restricted to $R_i$, the function $F_{\mathcal{N}, \theta} |_{R_i}$ can be realised as a one-layer NN with input neurons $n^{(1)}_i n^{(2)}_i d_i$ hidden neurons. By Proposition 1, $\mathcal{N}$ partitions $R_i$ into $\sum_{i=0}^{n^{(1)}_0 n^{(2)}_0 \times d_0} \binom{n^{(1)}_0 n^{(2)}_0}{d_0}$ distinct linear regions. Finally, we obtain

$$R_{\mathcal{N}} \leq R_{\mathcal{N}'} \sum_{i=0}^{n^{(1)}_0 n^{(2)}_0 \times d_0} \binom{n^{(1)}_0 n^{(2)}_0}{d_0} \sum_{i=0}^{n^{(1)}_L n^{(2)}_L \times d_L} \binom{n^{(1)}_L n^{(2)}_L}{d_L} \leq R_{\mathcal{N}''} \prod_{i=0}^{L} \sum_{i=0}^{n^{(1)}_0 n^{(2)}_0 \times d_0} \binom{n^{(1)}_i n^{(2)}_i}{d_i},$$

which completes the proof. \hfill $\square$

4. Calculation of the Number of Parameters for CNNs

Proof of Lemma 4 in the main paper. For the $l$-th layer, the $k$-th weight matrix $W_l^k$ has $f^{(1)}_l \times f^{(2)}_l \times d_{l-1}$ entries and there are $d_l$ such weight matrices. The bias vector has length $d_l$. Thus there are $f^{(1)}_l \times f^{(2)}_l \times d_{l-1} \times d_l + d_l$ parameters in the $l$-th hidden layer. Let $l$ range from 1 to $L$, the total number of parameters equals $\sum_{l=0}^{L} \left( f^{(1)}_l \times f^{(2)}_l \times d_{l-1} + d_l \right)$. \hfill $\square$

5. More Examples on the Maximal Number of Linear Regions for One-Layer ReLU CNNs

In this section, we list more examples on maximal number of linear regions for one-layer ReLU CNNs from Tables 1 to 5, which is calculated according to Theorem 2 in the main paper.
Table 1. The results for the maximal number of linear regions for a one-layer ReLU CNN with input dimension $2 \times 2 \times 1$, $d_1$ filters with dimension $1 \times 2 \times 1$, stride $s = 1$, and hidden layer dimension $2 \times 1 \times d_1$.

<table>
<thead>
<tr>
<th>$d_1$</th>
<th>$d_1 = 1$</th>
<th>$d_1 = 2$</th>
<th>$d_1 = 3$</th>
<th>$d_1 = 4$</th>
<th>$d_1 = 5$</th>
<th>$d_1 = 6$</th>
<th>$d_1 = 7$</th>
<th>$d_1 = 8$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R_N$ by Theorem 2</td>
<td>4</td>
<td>16</td>
<td>49</td>
<td>121</td>
<td>256</td>
<td>484</td>
<td>841</td>
<td>1369</td>
</tr>
<tr>
<td>Upper bounds by Theorem 1</td>
<td>4</td>
<td>16</td>
<td>57</td>
<td>163</td>
<td>386</td>
<td>794</td>
<td>1471</td>
<td>2517</td>
</tr>
<tr>
<td>Naive upper bounds</td>
<td>4</td>
<td>16</td>
<td>64</td>
<td>256</td>
<td>1024</td>
<td>4096</td>
<td>16384</td>
<td>65536</td>
</tr>
</tbody>
</table>

Table 2. The results for the maximal number of linear regions for a one-layer ReLU CNN with input dimension $1 \times 4 \times 1$, $d_1$ filters with dimension $1 \times 2 \times 1$, stride $s = 1$, and hidden layer dimension $1 \times 3 \times d_1$.

<table>
<thead>
<tr>
<th>$d_1$</th>
<th>$d_1 = 1$</th>
<th>$d_1 = 2$</th>
<th>$d_1 = 3$</th>
<th>$d_1 = 4$</th>
<th>$d_1 = 5$</th>
<th>$d_1 = 6$</th>
<th>$d_1 = 7$</th>
<th>$d_1 = 8$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R_N$ by Theorem 2</td>
<td>8</td>
<td>55</td>
<td>217</td>
<td>611</td>
<td>1396</td>
<td>2773</td>
<td>4985</td>
<td>8317</td>
</tr>
<tr>
<td>Upper bounds by Theorem 1</td>
<td>8</td>
<td>57</td>
<td>256</td>
<td>794</td>
<td>1941</td>
<td>4048</td>
<td>7547</td>
<td>12951</td>
</tr>
<tr>
<td>Naive upper bounds</td>
<td>8</td>
<td>64</td>
<td>512</td>
<td>4096</td>
<td>32768</td>
<td>2097152</td>
<td>16777216</td>
<td></td>
</tr>
</tbody>
</table>

Table 3. The results for the maximal number of linear regions for a one-layer ReLU CNN with input dimension $2 \times 3 \times 1$, $d_1$ filters with dimension $2 \times 2 \times 1$, stride $s = 1$, and hidden layer dimension $2 \times 1 \times d_1$.

<table>
<thead>
<tr>
<th>$d_1$</th>
<th>$d_1 = 1$</th>
<th>$d_1 = 2$</th>
<th>$d_1 = 3$</th>
<th>$d_1 = 4$</th>
<th>$d_1 = 5$</th>
<th>$d_1 = 6$</th>
<th>$d_1 = 7$</th>
<th>$d_1 = 8$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R_N$ by Theorem 2</td>
<td>4</td>
<td>16</td>
<td>64</td>
<td>247</td>
<td>836</td>
<td>2424</td>
<td>6126</td>
<td>13829</td>
</tr>
<tr>
<td>Upper bounds by Theorem 1</td>
<td>4</td>
<td>16</td>
<td>64</td>
<td>247</td>
<td>848</td>
<td>2510</td>
<td>6476</td>
<td>14893</td>
</tr>
<tr>
<td>Naive upper bounds</td>
<td>4</td>
<td>16</td>
<td>64</td>
<td>256</td>
<td>1024</td>
<td>4096</td>
<td>16384</td>
<td>65536</td>
</tr>
</tbody>
</table>

Table 4. The results for the maximal number of linear regions for a one-layer ReLU CNN with input dimension $6 \times 6 \times 1$, $d_1$ filters with dimension $1 \times 3 \times 1$, stride $s = 2$, and hidden layer dimension $3 \times 2 \times d_1$.

<table>
<thead>
<tr>
<th>$d_1$</th>
<th>$d_1 = 1$</th>
<th>$d_1 = 2$</th>
<th>$d_1 = 3$</th>
<th>$d_1 = 4$</th>
<th>$d_1 = 5$</th>
<th>$d_1 = 6$</th>
<th>$d_1 = 7$</th>
<th>$d_1 = 8$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R_N$ by Theorem 2</td>
<td>64</td>
<td>4096</td>
<td>250047</td>
<td>9129329</td>
<td>191102976</td>
<td>2537716544</td>
<td>23654622311</td>
<td>167557540697</td>
</tr>
<tr>
<td>Upper bounds by Theorem 1</td>
<td>64</td>
<td>4096</td>
<td>262144</td>
<td>16777216</td>
<td>1073741824</td>
<td>68719476736</td>
<td>4398045536122</td>
<td>28144369851217</td>
</tr>
<tr>
<td>Naive upper bounds</td>
<td>64</td>
<td>4096</td>
<td>262144</td>
<td>16777216</td>
<td>1073741824</td>
<td>68719476736</td>
<td>4398046511104</td>
<td>28147976710856</td>
</tr>
</tbody>
</table>

Table 5. The results for the maximal number of linear regions for a one-layer ReLU CNN with input dimension $3 \times 3 \times 2$, $d_1$ filters with dimension $2 \times 2 \times 2$, stride $s = 1$, and hidden layer dimension $2 \times 2 \times d_1$.

<table>
<thead>
<tr>
<th>$d_1$</th>
<th>$d_1 = 1$</th>
<th>$d_1 = 2$</th>
<th>$d_1 = 3$</th>
<th>$d_1 = 4$</th>
<th>$d_1 = 5$</th>
<th>$d_1 = 6$</th>
<th>$d_1 = 7$</th>
<th>$d_1 = 8$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R_N$ by Theorem 2</td>
<td>16</td>
<td>256</td>
<td>4096</td>
<td>65536</td>
<td>1048555</td>
<td>16721253</td>
<td>256376253</td>
<td>3459170397</td>
</tr>
<tr>
<td>Upper bounds by Theorem 1</td>
<td>16</td>
<td>256</td>
<td>4096</td>
<td>65536</td>
<td>1048555</td>
<td>16721761</td>
<td>256737233</td>
<td>3485182163</td>
</tr>
<tr>
<td>Naive upper bounds</td>
<td>16</td>
<td>256</td>
<td>4096</td>
<td>65536</td>
<td>1048576</td>
<td>16777216</td>
<td>268435456</td>
<td>4294967296</td>
</tr>
</tbody>
</table>
References


