







Modular Multimodal Machine Learning for Extraction of Theorems and Proofs in Long Scientific Documents

Shrey Mishra, Antoine Gauquier, Pierre Senellart

Motivation for Theorem KB

Volume of mathematical papers

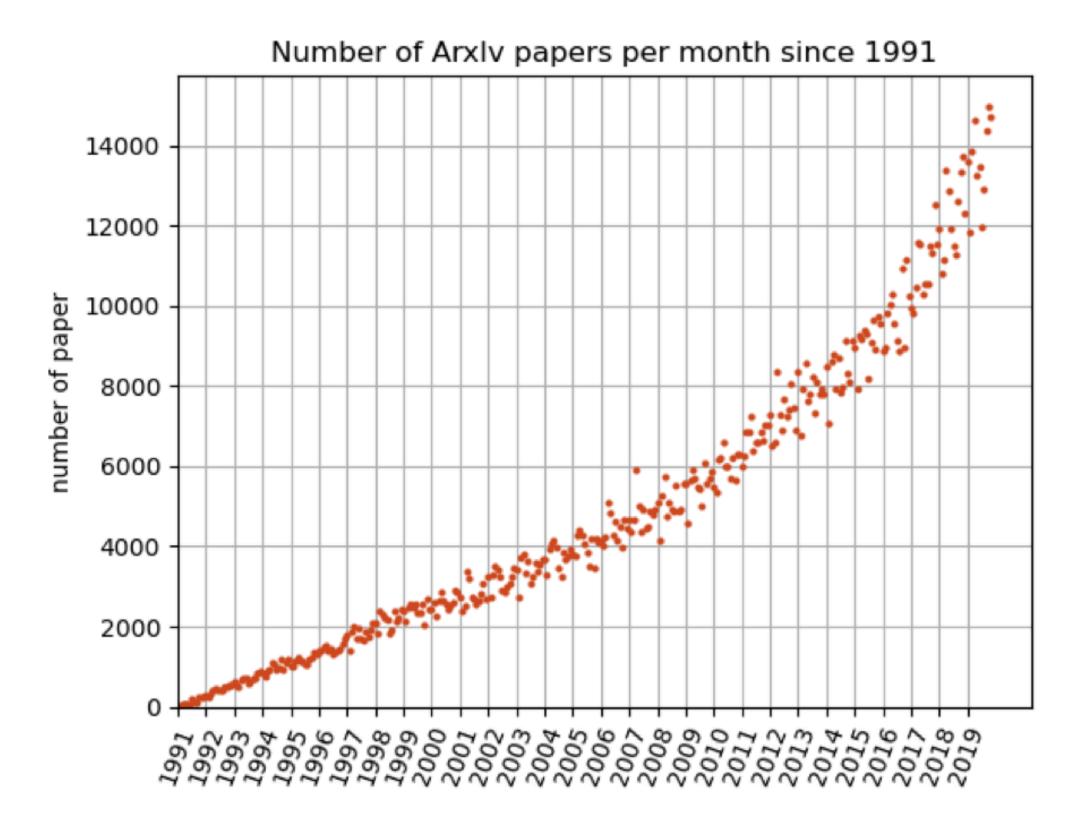


Figure 3.1: Total number of papers published on arXiv until 2019 [Del20]

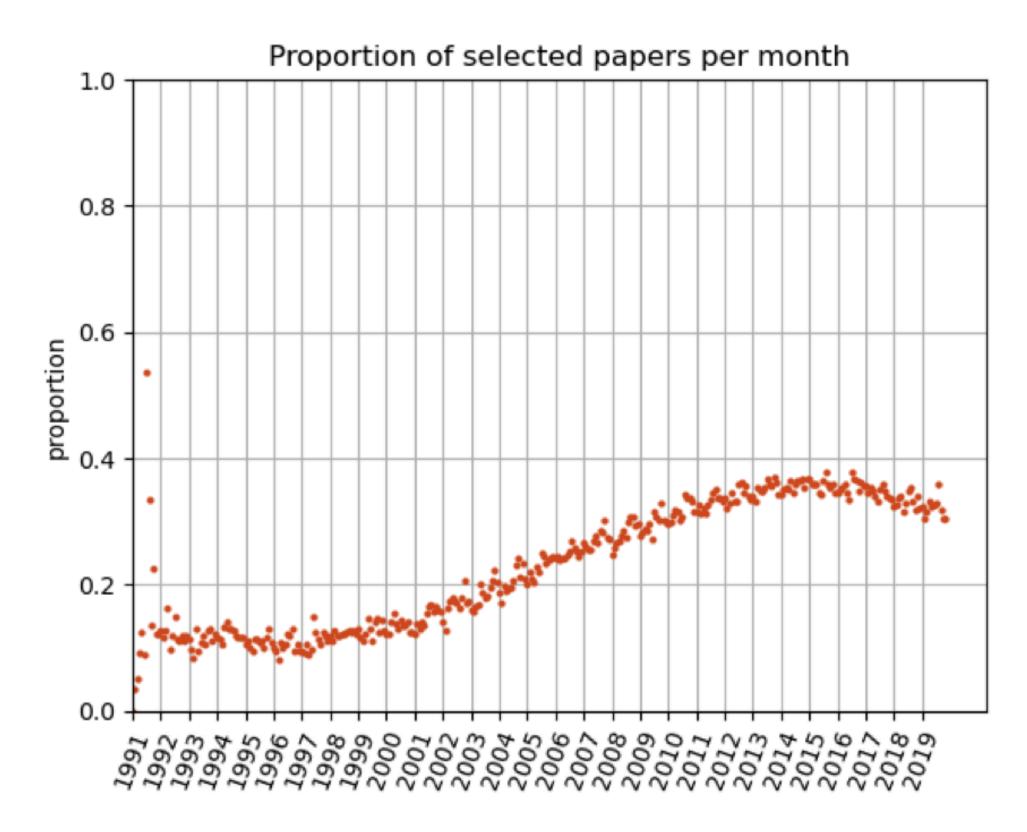


Figure 3.2: Papers with mathematical information such as Theorem, a Lemma or a Proposition [Del20]

Discovering existing knowledge

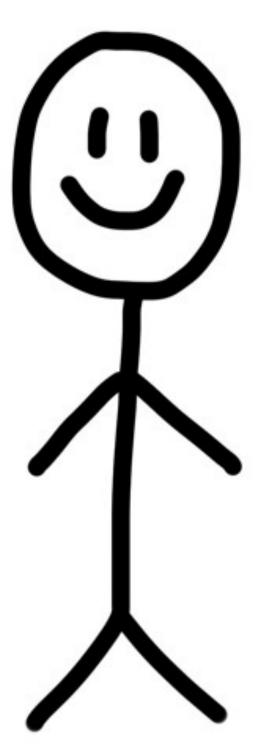






This other paper existed before

Michaël



(likes to read mathematical papers)

Hernich et al. 2012

Correcting publicised errors

Proof: Dichotomy for evaluating conjunctive queries



Nilesh et al. 2006

Proof v2: Dichotomy for evaluating conjunctive queries



Nilesh et al. 2012

Managing Complex Dependencies



Senellart et al. Theory of Computing systems, 2019

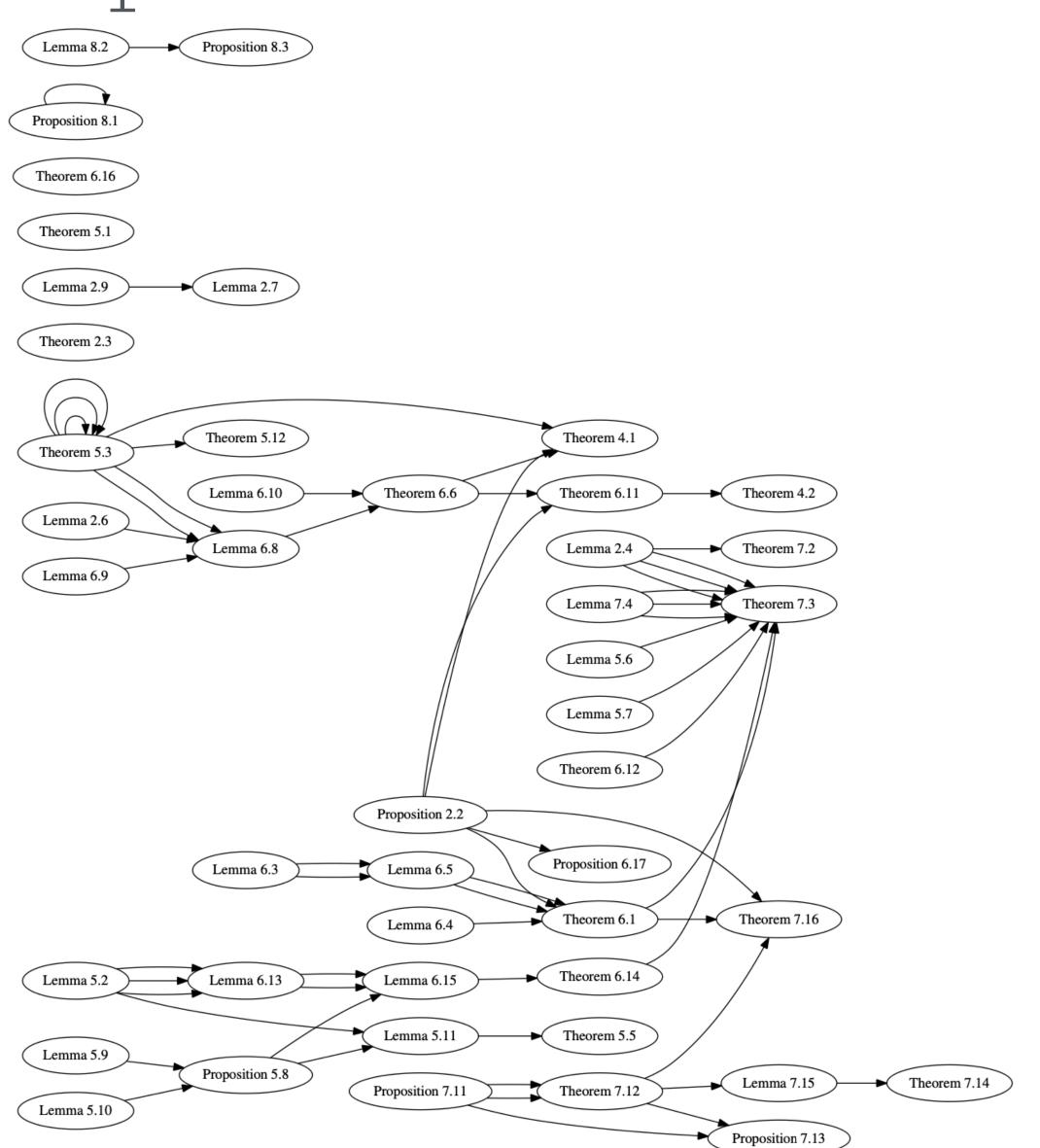


Table of contents

- Motivation of theoremkb
- Information Extraction task
- Unimodal and Multimodal backbones
- Sequential approach

Extraction task

High level overview of Theorem KB

PDF Dataset



Extracted results



Result graph

The Lagrangian remainder of Taylor's series, distinguishes $\mathcal{O}(f(x))$ time complexities to polynomials or not

Nikolaos P. Bakas, Elias Kosmatopoulos, Mihalis Nicolaou, Esavvas A. Chatzichristofis

Abstract

The purpose of this letter is to investigate the time complexity consequences of the truncated Taylor series, known as Taylor Polynomials [1–3]. In particular, it is demonstrated that the examinatio of the P = NP equality, is associated with the determination of whether the m^2 -derivative of a particular solution is bounded or not. Accordingly, in some cases, this is not true, and hence in

1 Univariate complexit

Definition 1. Let the given problem is a known analytic function f of one variable $x \in \mathbf{Z}^+$. Initially, the authors consider one-dimensional x, and later they generalize the results. Respectively, the time complexity of the given problem, according to the literature [4], may be written in the generic form

O(f(x)).

The execution time is usually calculated by some elementary algebraic operations of integers or real numbers [5], thus, this assumption is considered to be adequate and valid. Accordingly, the Taylor series expansion of f at $x + x_0$ may be written by

$$f(x) = f(x_0) + \frac{f'(x_0)}{1!}(x-x_0) + \frac{f''(x_0)}{2!}(x-x_0)^2 + \dots + \frac{f^{(n)}(x_0)}{n!}(x-x_0)^n + \dots ,$$
 with infinite terms, that is

 $f(x) = T_n(x) + R_n(x),$ (1) where $T_n(x)$ is the Taylor polynomial of order n, and $R_n(x)$ the <u>remainder</u> of the n^{th} degree Taylor

polynomial.

*Computation-based Science and Technology Research Center, The Cyprus Institute, 20 Konstantinon Kacufi Street, 2121, Aglantia Nicosia, Cyprus, e-mail: n.bakas@cyta.ccg;

*Democritus University of Thrace, Ricettal and Computer engineering, e-mail: locunatopities.duth.gr.

*Computation-based Science and Technology Research Center, The Cyprus Institute, 20 Konstantinon Kecufi Street, 2121, Aglantia Nicosia, Cyprus, e-mail: m.inclandowleys/inc.cy

*Intelligent Systems Lab & Department of Computer Science, Neapolis University Pafos, 2 Danais Avenue, 8042

*Pafos, Cyprus, e-mail: s-Antarichristiofistumpa.ccg.

Analysis on Real and Complex Manifolds

R. NARASIMHAN

University of Chicago Chicago, IL 60637, U.S.A.

N.H

NORTH-HOLLAND AMSTERDAM · NEW YORK · OXFORD **Definition 1.** Let the given problem is a known analytic function f of one variable $x \in \mathbb{Z}^+$. Initially, the authors consider one-dimensional x, and later they generalize the results. Respectively, the *time complexity* of the given problem, according to the literature [4], may be written in the generic form of:

O(f(x)).

Theorem 1. If $|f^{(n+1)}(x)| \le M$, the algorithm with O(f(x)) complexity, runs in polynomial time.

Lemma 1. If $|f^{(n+1)}(x)| > M$, f cannot be expressed as polynomials.

1.1.2 PRINCIPLE OF ANALYTIC CONTINUATION. If f is holomorphic (real analytic) in a connected open set $U(\Omega)$ in $C^n(R^n)$ and $D^a f(a) = 0$ for all $\alpha = (\alpha_1, \ldots, \alpha_n)$ and some $a \in U(\Omega)$, then $f \equiv 0$. In particular, if f vanishes on a non-empty open subset of $U(\Omega)$, then $f \equiv 0$.

1.1.3 WEIERSTRASS' THEOREM. If $\{f_r\}$ is a sequence of holomorphic functions, converging uniformly on compact subsets of U to a function f, then f is holomorphic in U. Moreover, for any α , $\{D^{\alpha}f_r\}$ converges to $D^{\alpha}f$, uniformly on compact sets.

1.2.6 COROLLARY. Let Ω be open in \mathbb{R}^n , X a closed subset of Ω and U an open subset of Ω containing X. Then, there exists a C^{∞} function ψ on Ω such that $\psi(x) = 1$ if $x \in X$, $\psi(x) = 0$ if $x \in \Omega - U$ and $0 \le \psi \le 1$ everywhere.

1.5.2 Lemma. Let $f \in C^{\infty}(\mathbb{R}^n)$ be m-flat at 0. Then, given $\varepsilon > 0$, there exists $g \in C^{\infty}(\mathbb{R}^n)$ which vanishes in a neighbourhood of 0 and such that

$$||g-f||_{m}^{R^{n}}<\varepsilon.$$

1.5.4 THEOREM OF BOREL. Given, for each *n*-tuple $\alpha = (\alpha_1, \ldots, \alpha_n)$ of non-negative integers, a real constant c_{α} , there exists an $f \in C^{\infty}(\mathbb{R}^n)$ such that

$$\frac{1}{\alpha!}D^{\alpha}f(0) = c_{\alpha}.$$

In other words, the mapping from $C^{\infty}(R^n)$ to the ring of formal powerseries in *n*-variables given by $f \mapsto T(f)$ is surjective.



1.2.6 COROLLARY. Let Ω be open in R^n , X a closed subset of Ω and U an open subset of Ω containing X. Then, there exists a C^{∞} function ψ on Ω such that $\psi(x) = 1$ if $x \in X$, $\psi(x) = 0$ if $x \in \Omega - U$ and $0 \le \psi \le 1$ everywhere.



1.5.2 Lemma. Let $f \in C^{\infty}(\mathbb{R}^n)$ be m-flat at 0. Then, given $\varepsilon > 0$, there exists $g \in C^{\infty}(\mathbb{R}^n)$ which vanishes in a neighbourhood of 0 and such that

$$\frac{||g-f||_m^{R^n}<\varepsilon.}{\uparrow}$$

1.5.4 THEOREM OF BOREL. Given, for each *n*-tuple $\alpha = (\alpha_1, \ldots, \alpha_n)$ of non-negative integers, a real constant c_{α} , there exists an $f \in C^{\infty}(\mathbb{R}^n)$ such that

$$\frac{1}{\alpha!}D^{\alpha}f(0)=c_{\alpha}.$$

In other words, the mapping from $C^{\infty}(\mathbb{R}^n)$ to the ring of formal power-series in *n*-variables given by $f \mapsto T(f)$ is surjective.



Lemma 1. If $| f^{(n+1)}(x) | > M$, f cannot be expressed as polynomials.

Proof. By utilizing Borel's theorem [8], stating that any formal series $\sum_{n=0}^{\infty} a_n (x-x_0)^n$ is the Taylor series of a C^{∞} -smooth function defined in an open neighborhood of x_0 , it is derived that if $f^{(n)}$ is not bounded, f cannot be written as a power series, and hence as a polynomial, thus the problem is not in \mathbf{P} . In other words, if the problem was in \mathbf{P} , it could be written as a polynomial, and this would be the Taylor series, which is absurd as no n exist such that the $f^{(n)}$ is limited by a M.

Theorem 1. If $\mid f^{(n+1)}(x) \mid \leq M$, the algorithm with $\mathcal{O}(f(x))$ complexity, runs in polynomial time.

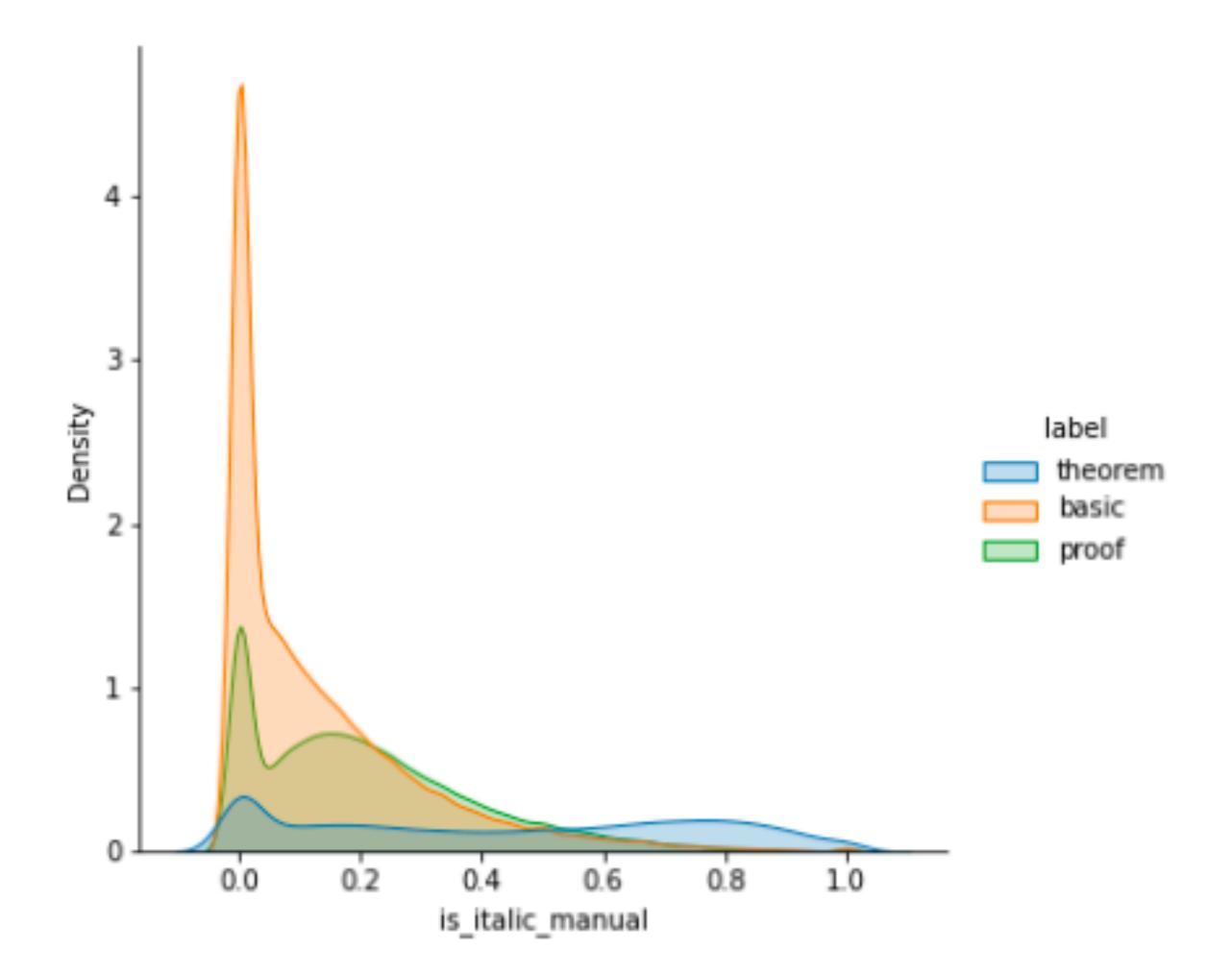
What is Extraction?

Theorem 1.8 ([HRVW09]). Let $f: \{0,1\}^n \to \{0,1\}^n$ be a one-way function, let X be uniformly distributed in $\{0,1\}^n$, and let (Y_1,\ldots,Y_m) be a partition of Y=f(X) into blocks of length $O(\log n)$. Then (Y_1,\ldots,Y_m,X) has next-block accessible entropy at most $n-\omega(\log n)$.

Proof. Since f is (t, ε) -one-way, the distributional search problem $(\Pi^f, f(X))$ where $\Pi^f = \{(f(x), x) : x \in \{0, 1\}^n\}$ is (t, ε) -hard. Clearly, (f(X), X) is supported on Π^f , so by applying Theorem 3.8, we have that $(\Pi^f, f(X), X)$ has witness hardness $(\Omega(t), \log(1/\varepsilon))$ in relative entropy and $(\Omega(t), \log(1/\varepsilon) - \log(2/\delta))$ in $\delta/2$ -min relative entropy. Thus, by Theorem 4.7 we have that $(Y_1, \ldots, Y_{n/\ell}, X)$ has next-block inaccessible relative entropy $(\Omega(t \cdot \Delta \cdot \ell^2/(n^2 \cdot 2^\ell)), \log(1/\varepsilon) - \Delta)$ and next-block inaccessible δ -min relative entropy $(\Omega(t \cdot \delta \cdot \Delta \cdot \ell^2/(n^2 \cdot 2^\ell)), \log(1/\varepsilon) - \log(2/\delta) - \Delta)$, and we conclude by Theorem 4.9.

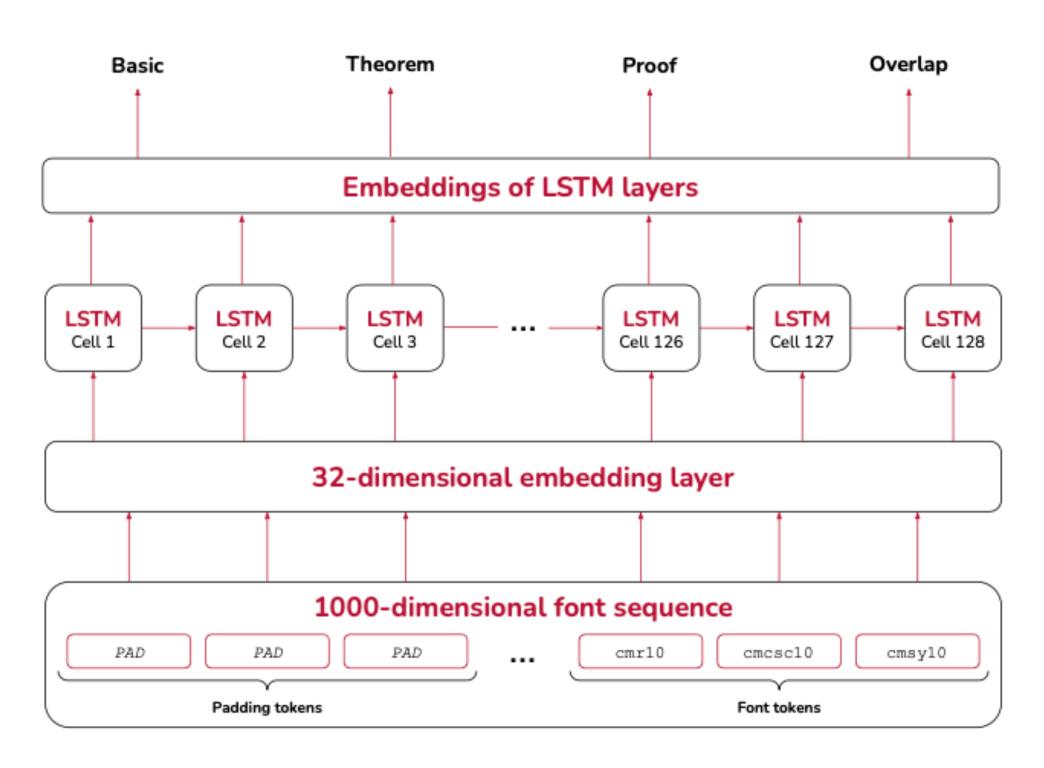
Unimodal backbones

Font based



Theorem 1.8 ([HRVW09]). Let $f : \{0,1\}^n \to \{0,1\}^n$ be a one-way function, let X be uniformly distributed in $\{0,1\}^n$, and let (Y_1,\ldots,Y_m) be a partition of Y=f(X) into blocks of length $O(\log n)$. Then (Y_1,\ldots,Y_m,X) has next-block accessible entropy at most $n-\omega(\log n)$.

Proof. Since f is (t,ε) -one-way, the distributional search problem $(\Pi^f, f(X))$ where $\Pi^f = \{(f(x),x): x \in \{0,1\}^n\}$ is (t,ε) -hard. Clearly, (f(X),X) is supported on Π^f , so by applying Theorem 3.8, we have that $(\Pi^f, f(X), X)$ has witness hardness $(\Omega(t), \log(1/\varepsilon))$ in relative entropy and $(\Omega(t), \log(1/\varepsilon) - \log(2/\delta))$ in $\delta/2$ -min relative entropy. Thus, by Theorem 4.7 we have that $(Y_1, \ldots, Y_{n/\ell}, X)$ has next-block inaccessible relative entropy $(\Omega(t \cdot \Delta \cdot \ell^2/(n^2 \cdot 2^\ell)), \log(1/\varepsilon) - \Delta)$ and next-block inaccessible δ -min relative entropy $(\Omega(t \cdot \delta \cdot \Delta \cdot \ell^2/(n^2 \cdot 2^\ell)), \log(1/\varepsilon) - \log(2/\delta) - \Delta)$, and we conclude by Theorem 4.9.



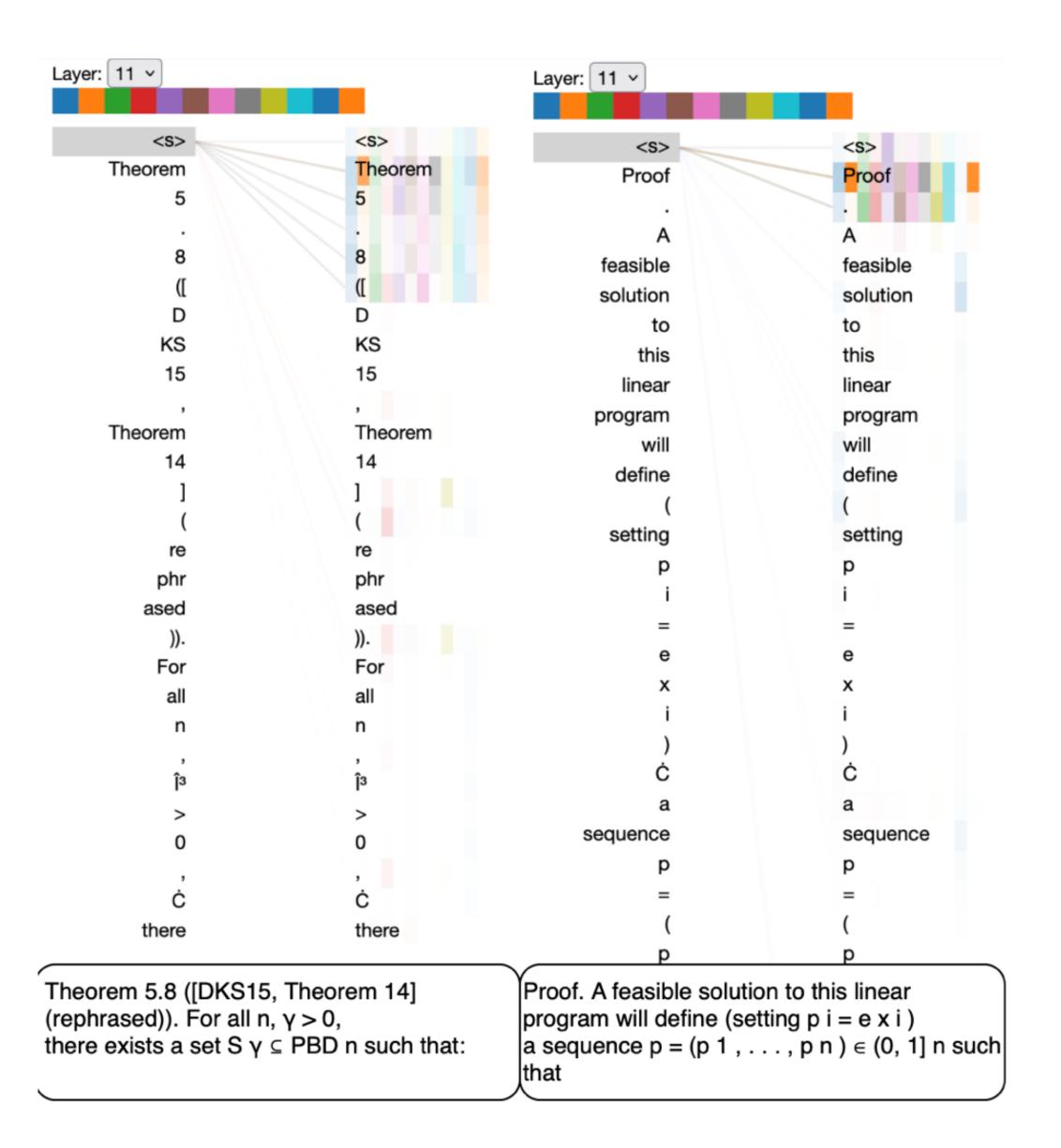
12

Vision Based

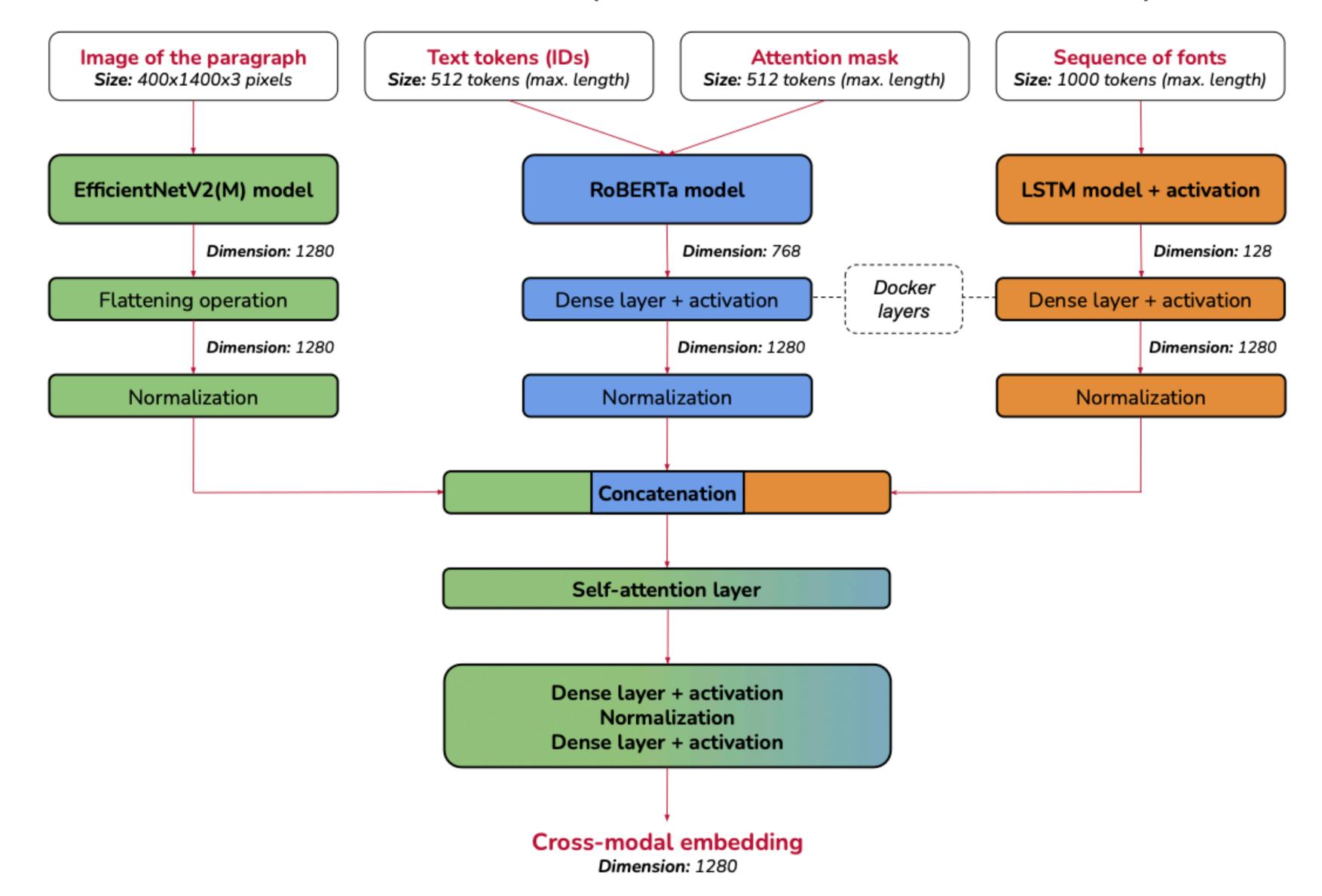
Proof. By Theorem 3, relative to a suitable oracle $A_{\mathcal{D}}$ (in fact, a random oracle suffices), there exists a signature scheme \mathcal{D} , such that any quantum chosen-message attack against \mathcal{D} must make superpolynomially many queries to $A_{\mathcal{D}}$. The oracle $A_{\mathcal{S}}$ will simply be a concatenation of $A_{\mathcal{M}}$ with $A_{\mathcal{D}}$. Relative to $A_{\mathcal{S}}$, we claim that the mini-scheme \mathcal{M} and signature scheme \mathcal{D} are both secure—and therefore, by Theorem 16, we can construct a secure public-key quantum money scheme \mathcal{S} .

Theorem 4. Suppose the ETH holds with constant c. Then for every $\alpha, \beta \in \mathbb{N}$ there exists a $\gamma = O(\alpha + \beta)$ such that

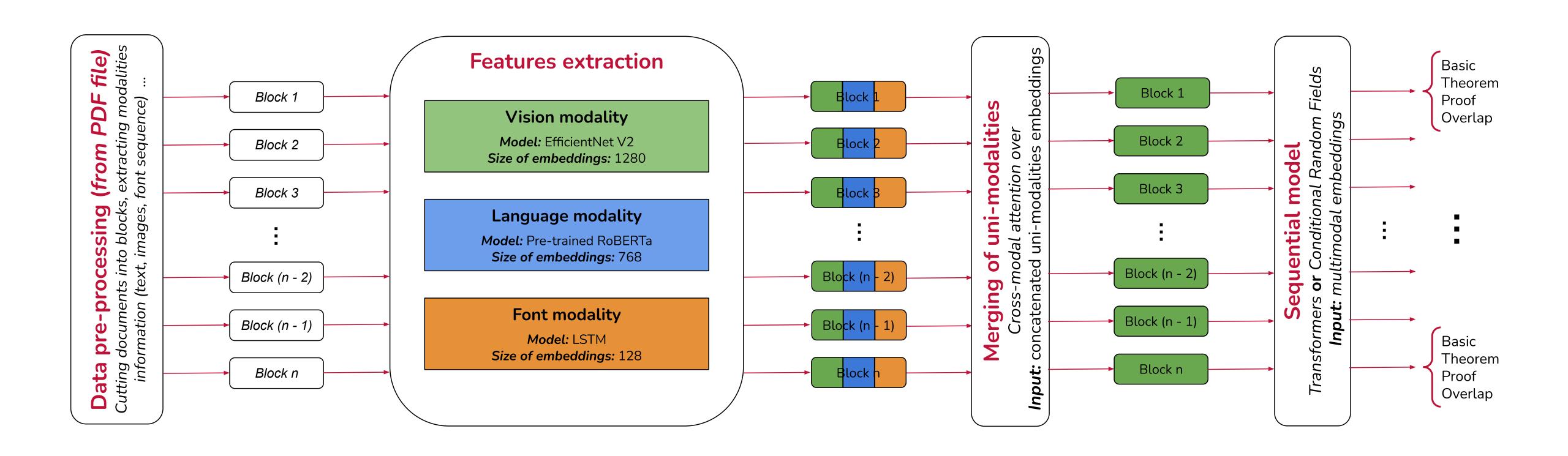
Text based

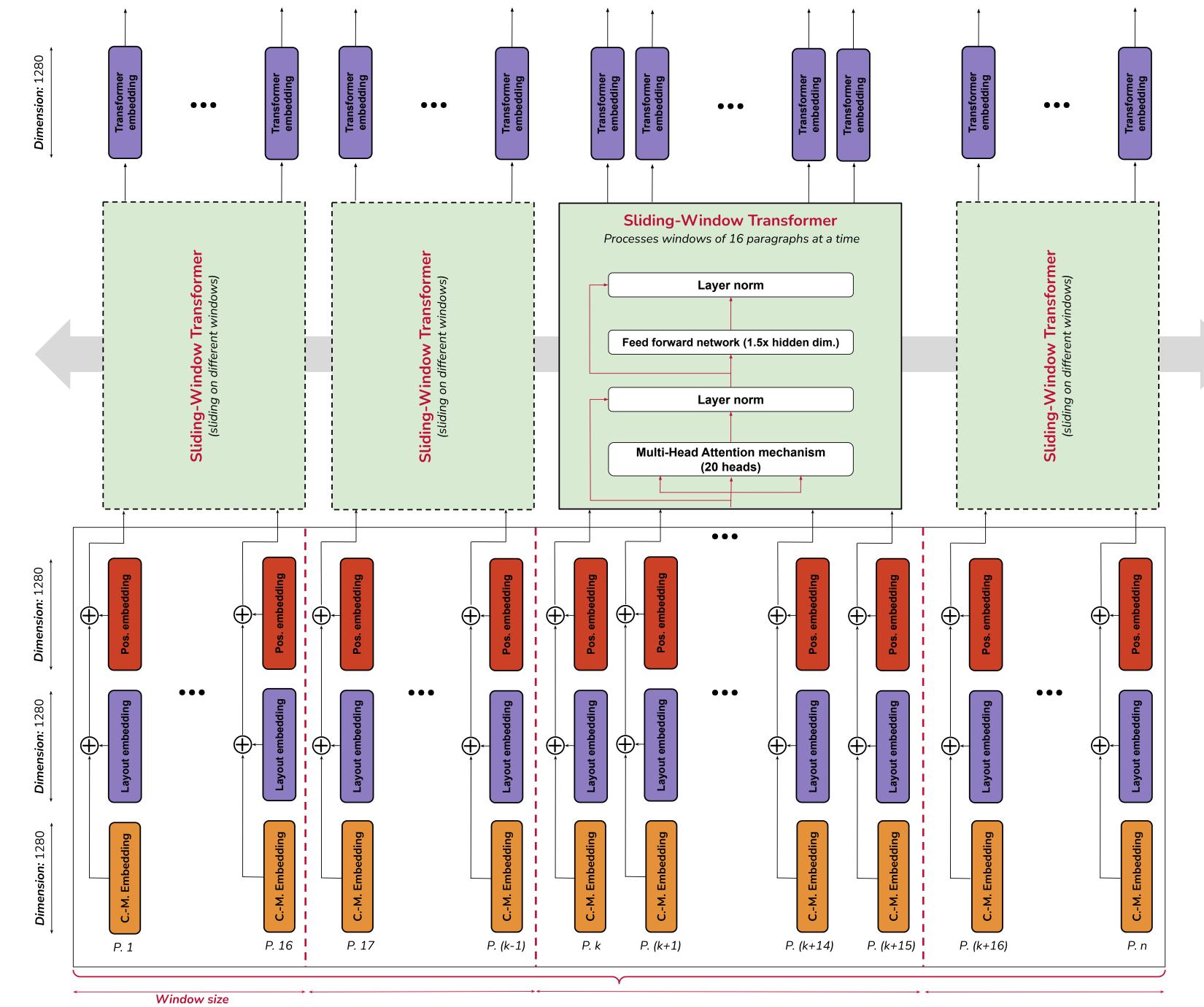


Multimodal model (raw features)



Sequential Approach





Performance of Sequence model

Modality	Model chosen	Seq. approach	#Batches	#Params (M)	Accuracy (%)	$\mathbf{Mean} \mathbf{F}_1 (\%)$
Dummy	always predicts $basic$				59.41	24.85
Top- k first word	use only first word				52.84	44.20
${\bf Line-based} [{\bf MPS21}]$	Bert (fine-tuned)			110	57.31	55.71

Closing remarks

- Our multimodal approach can be adapted to long documents
- Can make inference on entire pdf in a single forward pass
- Comparable, consistent and computationally efficient
- Unlike many other approaches that rely on an OCR preprocessing to be useful (LayoutLM) ours rely on Grobid which is many times faster
- Our approach is Fast (encoder only) and Scalable and applicable in real world (~200k pdfs tested)
- Our approach captures cross modality without adding special losses